NASA SP-68

SPACE TECHNOLOGY

Volume IV

SPACECRAFT GUIDANCE AND CONTROL

J. R. SCULL

Jet Propulsion Laboratory

Scientific and Technical Information Division
OFFICE OF TECHNOLOGY UTILIZATION 1966
NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
Washington, D.C.

FOR SALE BY THE SUPERINTENDENT OF DOCUMENTS, U.S. GOVERNMENT PRINTING OFFICE
WASHINGTON, D.C. 20402 - PRICE 55 CENTS

Preface

THIS IS THE FOURTH of a series of publications prepared as notes for a course in the Space Technology Summer Institute, given by the California Institute of Technology (Caltech) in cooperation with the Jet Propulsion Laboratory (JPL) from June 19 to July 31, 1964. The Institute—directed by E. E. Sechler, professor of aeronautics at Caltech—was sponsored by the National Aeronautics and Space Administration under Grant No. NsG–598 and was taught by engineers from industry, from Caltech, and from JPL. It is planned that the complete set will consist of—

Volume I	*Spacecraft Systems* by L. H. ABRAHAM DOUGLAS AIRCRAFT CO., INC.
Volume II	*Spacecraft Mechanical Engineering* by JAMES L. ADAMS, JPL
Volume III	*Spacecraft Propulsion* by F. E. MARBLE, CALTECH
Volume IV	*Spacecraft Guidance and Control* by J. R. SCULL, JPL
Volume V	*Telecommunications* by J. J. STIFFLER, JPL

Contents

CHAPTER 1

Spacecraft Guidance Philosophy

THE PRIMARY PURPOSE of this book is to discuss the guidance of lunar and planetary spacecraft with less-than-usual emphasis on satellites (ref. 1). The discussion will cover the tradeoffs among injection, midcourse, and terminal guidance as well as ways of mechanizing these systems by radio, inertial, or celestial techniques. Examples of these guidance systems as applied to some of the current lunar and planetary spacecraft will be described.

Spacecraft guidance is in many respects quite similar to guidance of a plane or ship with one major exception: guiding a typical spacecraft is similar to getting all of the speed needed for a transoceanic flight at the beginning of the flight, aiming in the direction of the target, and shutting off the engines with the control surfaces locked and coasting with zero drag to the target. The only exception to this type of approach is that used for electric-propulsion vehicles of very low thrust.

GUIDANCE ERRORS

For successful space missions, the permissible guidance errors are relatively small. For a satellite in a 300-mile circular orbit, an error of 1° in the elevation angle of the velocity vector would cause an apogee-perigee error of 140 miles. A corresponding error caused by a velocity difference of 1 ft/sec from the nominal would be 0.7 mile.

In a lunar trajectory, the errors are much larger. For example, a 1° error in the velocity vector would cause a miss distance of 2500 miles. A velocity error of 1 ft/sec would cause a miss distance of approximately 20 to 1000 miles, depending upon the speed of the trajectory, with fast trajectories normally causing somewhat smaller effects of a given error than those of low speed.

To improve the mission accuracy or, in some cases, to make possible flights which otherwise could not be performed at all, some form of midcourse or terminal guidance is usually required. One might ask: when should midcourse or terminal guidance be used rather than just improving the injection guidance system? The answer is that this is largely a matter of cost—not only in dollars but in some cases additional weight or additional complexity.

For example, the improvement of injection guidance accuracies to much beyond 4 or 5 ft/sec, or 0.1° of the velocity vector, usually turns out to be relatively costly. In a typical satellite having injection velocities of 25 000 ft/sec, achieving a velocity accuracy of 2.5 ft/sec would require a velocity-measuring device—either an accelerometer or Doppler device—accurate to 0.01 percent. A relatively simple midcourse guidance system accurate to only 1 percent, using amounts of corrective propellants of only 2 to 5 percent of the overall payload, could reduce the target error by a factor of as much as 30. Thus, with relatively available components and techniques, lunar errors of 50 miles or near-planet errors of 5000 miles might be obtained.

If a more precise approach to the target is required, some form of terminal guidance is necessary. A planetary miss of perhaps 5000 miles, added to the imprecision in our knowledge of the astronomical unit, could amount to 20 000 miles at Mars or Venus. The application of a relatively small propulsion system with a simple guidance system of the same level described for midcourse permits, by vernier means, reduction of this error by a factor of approximately another 30 to 50.

There is a tradeoff involved in selecting the point at which the midcourse or terminal correction should be made. The longer one waits to make the maneuver, either midcourse or terminal, normally the more accurate the data become. In addition, as the vehicle gets farther away from the launch point, the target error resulting from a given velocity or angular error in the maneuver is reduced. However, as the target is approached, more fuel is required to make this correction.

TYPES OF GUIDANCE SYSTEMS

There are a number of guidance systems applicable to injection, midcourse, and terminal guidance. As with aircraft or ship navigation, certain guidance mechanizations appear to be more applicable to specific phases of a space journey than others. For satellites, as an example, almost any form of injection guidance appears to be applicable. Radio, inertial, and optical aids or a combination of the three are useful for direct-ascent trajectories where the point of burnout of the rocket is within sight of the launcher or over some definite location. However, if coasting trajectories—which will be described later—are used, methods which depend on radio guidance are usually less applicable because the point of injection, where the greatest accuracy is needed, is not normally within sight of the launcher. For many lunar and planetary missions, injection does not occur over any uniquely located spot on the globe. As an example, for the 24-hour communica-

tions satellite, the vehicle is required in some instances to coast for as long as 16 hours before the final increment of burning takes place. Depending on the trajectory and launch point, it may occur over almost any point on the Earth. It would be difficult to have a radio guidance system set up to handle this. When a coasting orbit is called for, some form of inertial guidance or a combination of inertial guidance and optical references, such as horizon scanners, is normally required.

The trajectory of the Ranger spacecraft is shown in figure 1.1. The trajectory shown is relatively a low-energy one—actually a high-ellipticity satellite orbit around the Earth. A closeup of this trajectory relative to the Earth (fig. 1.2) shows that, for a launch from the Atlantic Missile Range, it may not always be possible to have the desired point of injection (the perigee of the orbit) occur within sight of the range or over any specific location. As the time of day and the lunar month change, the desired injection point may vary over practically the entire Atlantic Ocean. Figure 1.3 shows a parking orbit for a Ranger lunar flight with the injection point 5000 miles downrange and thus out of sight of the launcher. For most lunar and planetary vehicles, the optimum flight uses a coasting trajectory and thus requires inertial rather than radio guidance.

A coplanar representation of some typical trajectories to Venus and Mars in 1962 is shown in figure 1.4. The near-Earth hyperbolic portion of the 1962 Mars trajectory is illustrated as an example in figures 1.5 and 1.6. Since the locus of perigee points (approximately the injection loci) occurs either in the Southern Hemisphere or in low northern latitudes, the only method of achieving the proper trajectory is to launch from the Southern Hemisphere or to use a coasting (parking) orbit of nearly halfway around the Earth. Again, because of the rotation of the Earth around its axis and around the Sun, the exact point of injection may occur over a very large area. The most practical way of solving this injection problem is to use inertial techniques.

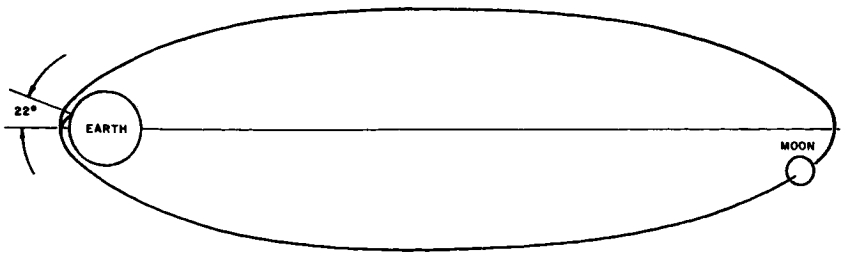


FIGURE 1.1—Ranger-A lunar trajectory, in plane of trajectory.

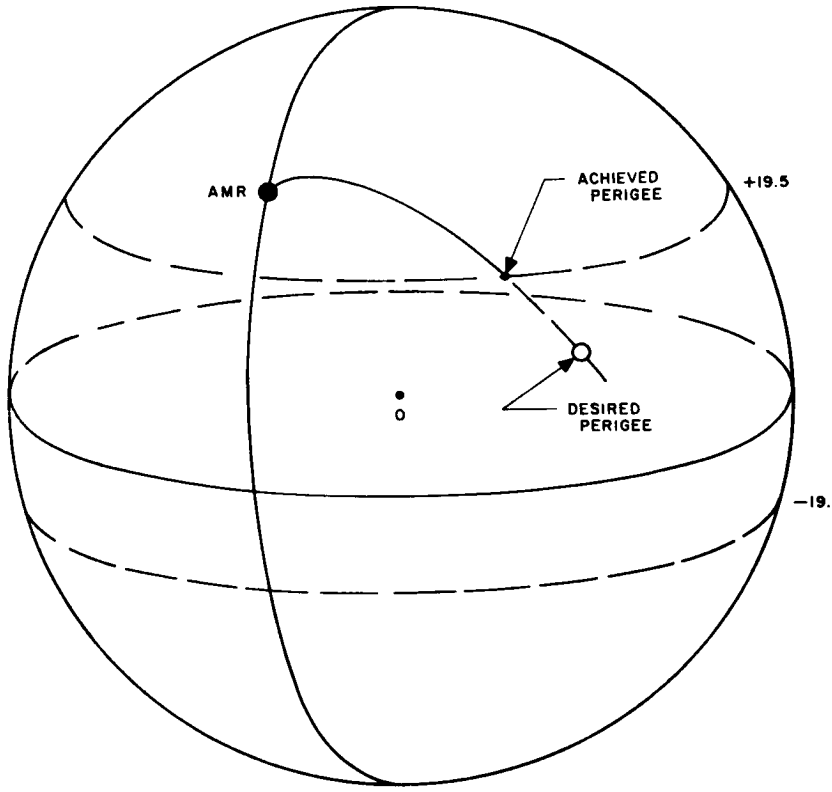


FIGURE 1.2—Ranger-A geocentric trajectory geometry.

MIDCOURSE GUIDANCE

For midcourse guidance, one of the main advantages is that there is a fair amount of time in which to determine what corrections are required and to smooth the data. Typically, the accuracy required for midcourse guidance is similar to that for injection as far as the actual numerical values of the angular and velocity measurements are concerned. The main difference lies in the fact that just before

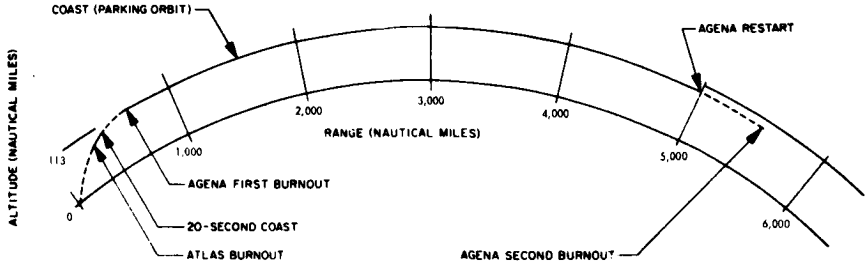


FIGURE 1.3—Ranger-A lunar parking orbit.

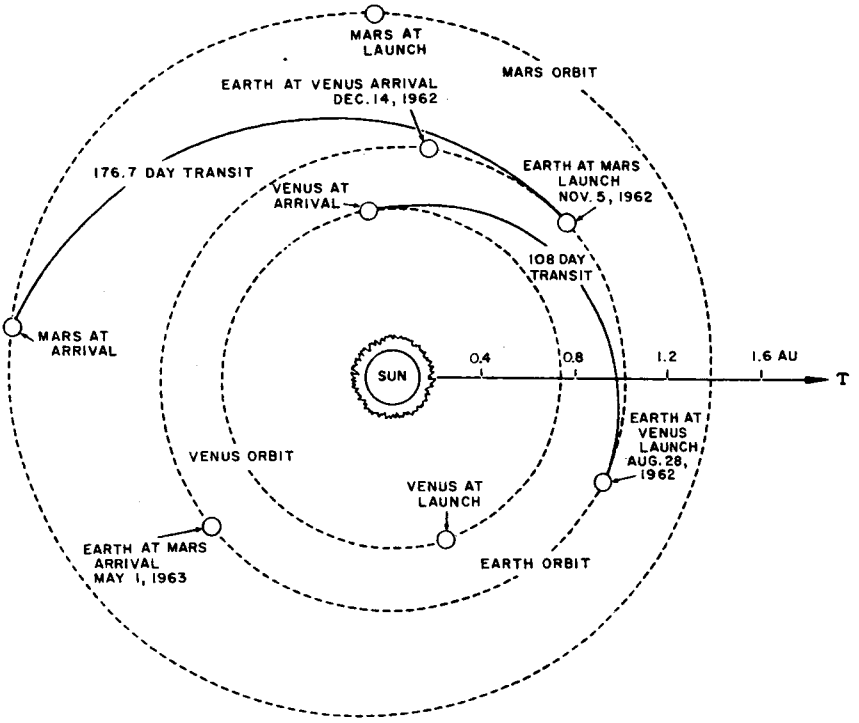


FIGURE 1.4—1962 heliocentric transfer.

the midcourse maneuver, the measurement of the required velocity-vector change can be made over a considerable length of time during which the spacecraft is subjected only to the slowly varying gravitational field which can easily be taken into consideration. In addition, since the midcourse velocity increment is relatively small compared with the total velocity, the required precision of the maneuver or its measurement is accordingly reduced.

Inertial guidance does not appear to be well suited for the determination of the required midcourse maneuver. Since an inertial system has errors proportional to time, its accuracy degrades after injection even though the environment is improved. A low-accuracy inertial system can be used, however, to orient the thrust of the midcourse rocket and to terminate the thrust after the proper velocity increment has been obtained.

Under these circumstances, the system that appears to be the most attractive for unmanned flights is a radio midcourse-guidance system based on ground measurements from a worldwide tracking network, use of a reasonable amount of time to smooth the data, and transmission of a command to the spacecraft to make a given correction.

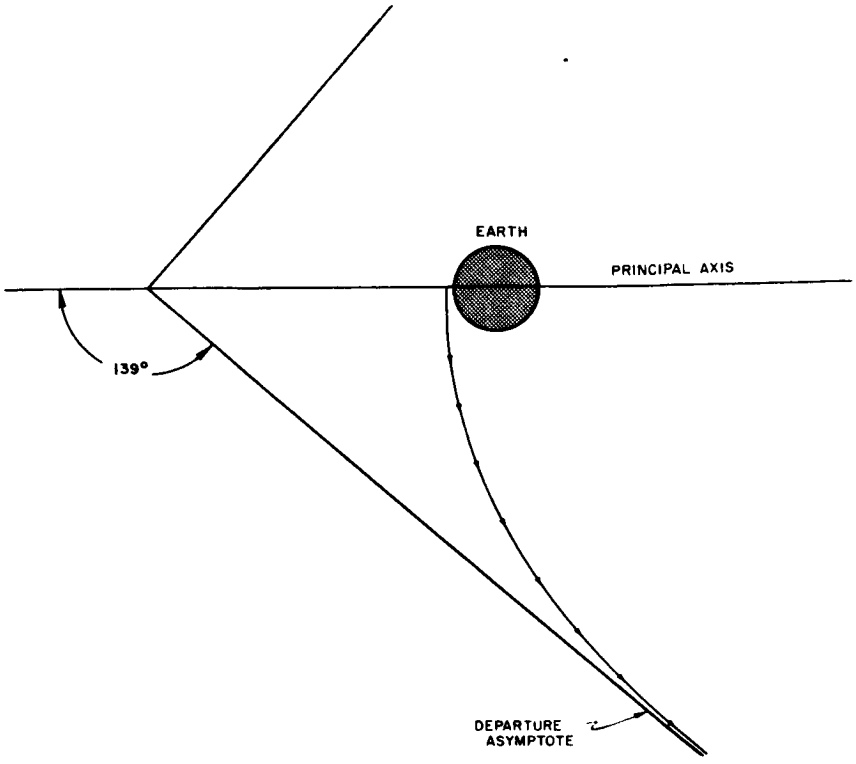


FIGURE 1.5—Geocentric departure hyperbola.

After the correction, the velocity changes achieved can be determined and, if necessary, the process can be repeated one or two times. Normally, only one midcourse maneuver is required. This is discussed further in chapter 11.

Midcourse celestial measurements also have an application in planetary trajectories. Usually, however, they are used only when the spacecraft is very far from Earth or when the trajectory is a long distance from Earth, as in the case of flights to planets outside the orbit of Mars. In general, radio guidance systems are useful near Earth; celestial guidance systems for midcourse are useful farther away from Earth. Celestial guidance is discussed in chapter 12.

TERMINAL GUIDANCE

For terminal-guidance purposes, some form of spacecraft-contained equipment is normally required. As an example, a typical celestial fix for terminal guidance can be obtained by measuring a minimum of three angles, usually between two stars and the destination planet;

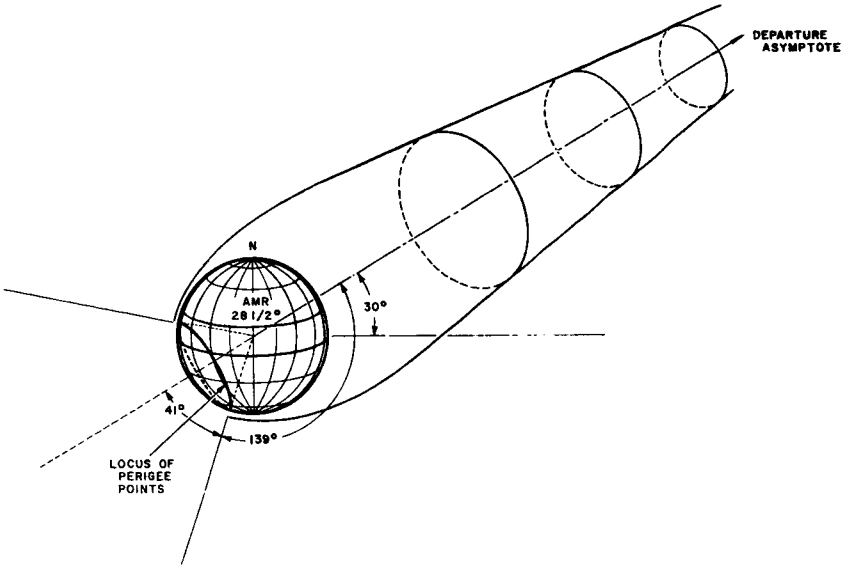


FIGURE 1.6—Departure geometry (Mars 1962).

one of the stars may be the Sun. One or two corrections at about a million miles from a destination planet would again reduce the target errors due to midcourse guidance by a factor of 30 to 50 and would allow accuracies of a few hundred miles using about 5 percent of the overall spacecraft payload weight for fuel to make the corrective maneuver. This is discussed further in chapter 14.

For an actual soft landing on a planet or the Moon, some method of range and velocity information is necessary in addition to celestial measurements. Quite often the optimum method for this is by Doppler or some other form of radar. Chapter 13 describes a lunar landing technique presently under development.

CHAPTER 2

Optical Sensors

EARLY BACKGROUND

ALTHOUGH THE ATTITUDE of early spin-stabilized spacecraft could not be controlled once it had been established, some scientific experiments required knowledge of the instantaneous attitude to achieve mission success. The Pioneer I space probe, which on October 11, 1958, reached an altitude of 70 600 miles from Earth, contained an optical scanner fixed to the spin-stabilized spacecraft. This scanner consisted of a lens with a photocell at its focus. As the spacecraft rotated at 2 rps, the intensity of any bright object seen by the scanner would be telemetered back to Earth. The resolution of the system was 0.5° by 0.5°, and it could provide a crude TV-type picture as well as attitude data. However, Pioneer I did not provide any useful optical data because its trajectory did not permit it to pass close enough to the Moon.

The Vanguard II satellite, launched on February 17, 1959, contained a pair of similar spin-scan detectors located in diametrically opposite positions at 45° from the spin axis. The output of the photocells at the focus of the optical telescopes was recorded on magnetic tape and was read out once per orbit by a ground station. Variations in the light reflected from clouds and Earth were obtained, but useful TV-type pictures or attitude information were extremely difficult to obtain because of precessional motion of the spinning spacecraft.

Pioneers III and IV, launched on December 6, 1958, and March 3, 1959, contained a pistol-shaped light sensor consisting of a lens and two photocells behind small apertures located at the focal plane. The apertures were spaced so that only a comparatively large image would illuminate both photocells simultaneously. The purpose of this detector was to indicate the proximity of the Moon as the spinning payload passed nearby. The ultimate intended use for the sensor was to trip the shutter of a camera which would relay pictures of the Moon back to Earth. The sensor was designed to operate when the spacecraft was within 22 000 miles of the Moon. Since the nearest approach of Pioneer IV was 37 300 miles, no useful data were obtained.

The first spacecraft to carry a three-axis attitude-control system was Discoverer I, launched on February 28, 1959. Although this

first system tumbled because of stability problems, the Discoverer II satellite, launched on April 13, 1959, was successfully attitude stabilized. The Discoverer control system uses an infrared horizon scanner as an attitude reference for two of its axes.

On October 4, 1959, the Soviet spacecraft Luna III was launched on a lunar trajectory. The spacecraft was spin stabilized until it reached the vicinity of the Moon. As it reached a predetermined position, it stopped spinning and oriented itself optically on a line between the Sun and the Moon. The third axis was apparently not optically stabilized, since photographs taken show a slowly varying roll attitude. On completion of the photography, the spacecraft resumed spinning for proper temperature control. Little information has been disclosed on the details of the sensors or the control system.

Pioneer V, which was launched toward the orbit of Venus on March 11, 1960, contained a photoelectric aspect sensor. The purpose of this sensor was to determine the direction toward the Sun during each revolution of the spin-stabilized spacecraft. This information was to be used in determining the direction of the magnetic field measured by a magnetometer. Although the spacecraft was quite successful and set the record for long-distance telemetry, the optical aspect sensor failed during launch and provided no useful information.

The Explorer X space probe, launched on March 25, 1961, contained an optical aspect sensor to determine the instantaneous orientation of the spin-stabilized spacecraft. The sensor consisted of a lens and a pair of slits at the focal plane with a photocell behind the slits. The sensor had a relatively large field of view, and as the image of the Sun, Earth, or Moon passed over the slits, a signal was telemetered to Earth. The duration and spacing between the signals was interpreted on the ground to give information on the probe attitude and, with less accuracy, its position in the trajectory. This information was used to provide directional information for mapping the magnetic fields of the Sun and Earth by means of a rubidium-vapor magnetometer.

Many Earth satellites use optical attitude sensors and are mentioned here briefly. The Tiros meteorological satellites contained six solar-aspect sensors to determine pointing direction for proper orientation of the pictures. Explorer VII contained radiation-balance, solar-ultraviolet, and X-ray detectors, which provided directional information in addition to scientific data. The Orbiting Solar Observatory is stabilized in space by Sun sensors. The Mercury spacecraft derived its attitude-control information for two axes from an infrared horizon scanner in the automatic control mode. The Nimbus meteorological satellite and many others use an infrared horizon scanner for control of two axes, with a Sun sensor or gyroscope for the third axis. At this time, the system of highest accuracy is contained in the Orbiting Astro-

nomical Observatory, which will use a separate star sensor for rough alinement and the main 36-inch Cassegrain telescope for the final alinement, with an ultimate accuracy of 0.1 second of arc.

PRESENT TECHNIQUES

The increased load-carrying capacity of launch vehicles has now permitted lunar and planetary spacecraft to be continuously attitude stabilized throughout their mission lifetimes, and this, in turn, has allowed a significant improvement in many of the subsystems of the spacecraft. As examples, telemetry bandwidth is increased because a high-gain directional antenna can be used, and the weight of the solar power system can be reduced by a factor of 6; scientific experiments, otherwise difficult or impossible, now become relatively easy.

The Sun and the Earth are used as attitude references for the Ranger lunar probe. The Sun is chosen as the reference for two of the spacecraft control axes because it is extremely easy to identify with simple detectors and because temperature control of the critical spacecraft components is thereby simplified. The third degree of freedom is controlled to orient one axis of a directional antenna toward the Earth. These reference directions are used as starting points for any angular maneuvers or measurements. Additional celestial references may be required in the future for determining approach or terminal guidance maneuvers in the vicinity of the destination planet for planetary orbits or landings.

For trajectories away from the Sun, such as to Mars or Jupiter, the Earth becomes a poor third-axis reference. Its luminous intensity is low because it is seen as a crescent, and it is difficult to track because, during part of any trajectory away from the Sun, the Sun falls in the field of view of the Earth sensor at intensities of 12 orders of magnitude greater. Thus, for trajectories away from the Sun, some other celestial object is preferable, such as the star Canopus, which is approximately 90° away from the sunline.

The Sun sensors used in the Ranger spacecraft are shown schematically in figure 2.1. These sensors are optical devices which use a shadow vane to control the relative amount of illumination on cadmium sulfide photoconductor cells. The primary cells, F and A, are located on the side of the spacecraft toward the Sun. These primary cells are connected in bridge fashion, with additional secondary cells located around the spacecraft to complete a spherical field of view. The characteristics of the primary cells in a bridge are

$$E_0=\frac{E(R_F-R_A)}{R_A+R_F+(R_AR_F/R_L)} \tag{2.1}$$

238-635 O—67——2

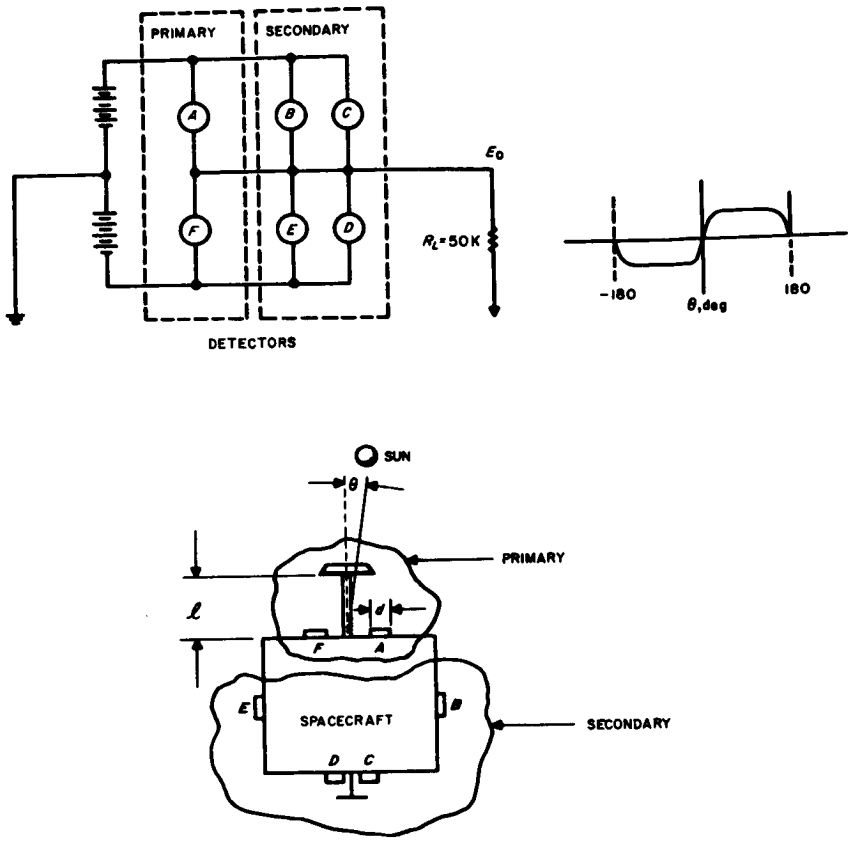


FIGURE 2.1—Sun sensor schematic.

The resistance of the cell varies as a function of incident angle approximately by

$$R_{A,F}=Ae^{\pm b\Delta\theta} \tag{2.2}$$

Using these cells in the bridge connection, a typical output of 10 V/deg is obtained across the load resistor. The slope of the detector at null may be determined by combining equations (2.1) and (2.2) and differentiating. The result is

$$\left.\frac{\partial E_0}{\partial \Delta\theta}\right|_{\Delta\theta=0}=\frac{2Eb}{(A/R_L)+2} \tag{2.3}$$

Additional Sun sensors, called secondary or Sun-finder cells, are located around the spacecraft to provide a spherical field of view for the entire group of detectors. Two complete sets of primary and secondary sensors, one each for the pitch and yaw axes, are required.

The null stability of the primary Sun sensor is about 0.02° and its resolution is 5 seconds of arc.

The Ranger Earth sensor is similar in principle to the primary Sun sensor, except that the low level of light from the Earth requires more sensitive detectors. The cadmium sulfide solid-state photoconductors are replaced with photomultiplier tubes. Instead of using two independent sensor systems to obtain two-degree-of-freedom information for the Earth sensor, these functions are combined using only three detectors.

A schematic of the Ranger Earth sensor is shown in figure 2.2. This sensor also uses a shadow mask in a slightly different configuration. The shadow mask partially obscures the cathodes of three Du Mont K2103 photomultiplier tubes. The outputs of the photomultipliers are summed and differenced much like a monopulse radar to provide left-right and up-down error signals and an AGC signal for changing the operating point of the tubes as a function of the light level. The Earth sensor has a 40° by 60° field of view, and the linear range is $\pm 2.5°$, relatively independent of target size for targets smaller than 5°. The null stability of this sensor is about $\pm 0.2°$. The sensitivity of the Earth sensor covers the range from near the Earth to about 2 million miles, or from approximately 20 ft-c to 0.01 ft-c.

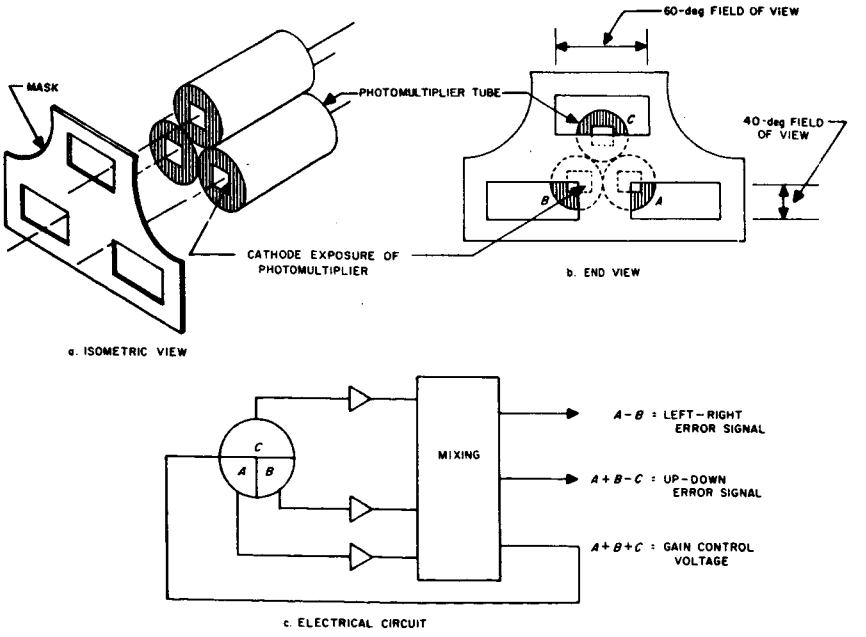


FIGURE 2.2—Earth sensor schematic.

The flight test results of the Ranger attitude sensors have been relatively successful. They have been successfully oriented by their Sun and Earth sensors to within ⅓° of the sensor null throughout the long "cruise" portion of the trajectory between the Earth and the Moon.

Planetary missions require attitude sensing to distances considerably greater than the threshold of the Earth sensor will allow. To meet this requirement, other sensors have been developed. In general, the sensitivity of the Earth sensor is limited by the small light-collecting area of the photomultiplier. The same photomultiplier combined with light-collecting optics will allow tracking of the Earth to distances beyond the orbit of Venus. The dynamic range of a long-range Earth sensor designed on this principle is from 2.2×10^{-7} W/cm^2 to 2.5×10^{-10} W/cm^2, which will allow orientation out to 60 million kilometers from the Earth. The long-range Earth sensor consists of an $f/1.2$, 50-millimeter collecting lens, a vibrating-reed chopper, and a K2103 photomultiplier. The aperture in the vibrating-reed chopper is shaped to produce digital error signals from the photomultiplier for both the roll and hinge directions, where "hinge" is defined as the axis, perpendicular to roll, used for orienting the high-gain antenna relative to the spacecraft. The field of view of the long-range Earth sensor is 4° by 10° and its null stability is about 0.1°.

Since the entire spacecraft is oriented in a fixed attitude with respect to the Sun and Earth, scientific instruments which must be directed toward the Moon or a planet require additional degrees of freedom relative to the spacecraft. An example of an infrared horizon scanner to orient scientific instruments toward a planet while on a flyby trajectory is shown in figure 2.3. The atmospheric temperatures of Venus and Mars range from 205° to 280° K, and this discontinuity between atmosphere and space background is detectable by an infrared scanner. This scanner operates on planet-emitted radiation of between 5-μ and 15-μ wavelength, using germanium optics and an immersed thermistor-bolometer detector. Scanning is provided by a pair of germanium wedges counterrotating at different speeds to produce a four-leaf rosette pattern. Time-coded error signals are detected and applied to the inputs of a servo system, which orients the scanner and associated scientific instruments toward the planet. The horizon scanner has a field of view of 70° and is capable of operating over a range of about 6000 to 160 000 kilometers from Venus or Mars with a null uncertainty of 0.1°. Although the scanner field of view is 70°, the instantaneous field of view seen by the detector is only 0.5° by 0.5°.

For more advanced lunar and planetary missions, more sophisticated attitude sensors are required. Already developed are star trackers

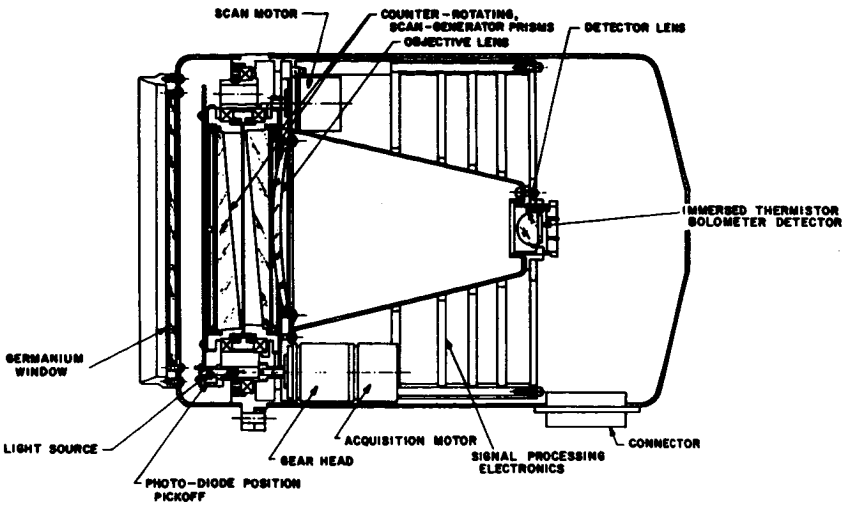


FIGURE 2.3—Horizon scanner.

which will recognize, identify, and track the star Canopus. Other advanced requirements include Sun, star, and planet trackers to measure angular separation between celestial bodies to an accuracy of 2 to 10 arc-sec for navigational fixes during a lunar or planetary trajectory.

TEST EQUIPMENT

The environment of outer space imposes several unique requirements for simulation and test equipment. Trackers which must identify specific stars must be tested with a simulation of the intensity and spectrum of the star as seen above the Earth's atmosphere. Sun sensors used for planetary spacecraft must be calibrated over a large range of solar intensity and apparent diameter to make sure they will function over the entire trajectory. The only way that an Earth sensor or a horizon scanner may be tested prior to flight is with a simulator. The thermal radiation balance of a spacecraft when it is bathed in sunlight but radiating to nearly absolute zero in the directions away from the Sun must also be determined before flight. A few of the more interesting examples of special optical simulators and test equipment to meet these requirements for lunar and planetary spacecraft are described in the following paragraphs.

A simulator (fig. 2.4) has been designed and built to provide an image of a planet illuminated by the Sun. Apparent size, phase of illumination, albedo, and number of moons can be varied to suit the conditions being simulated. The image of the planet or planet-moon combination is collimated so that it appears to be focused at an infinite

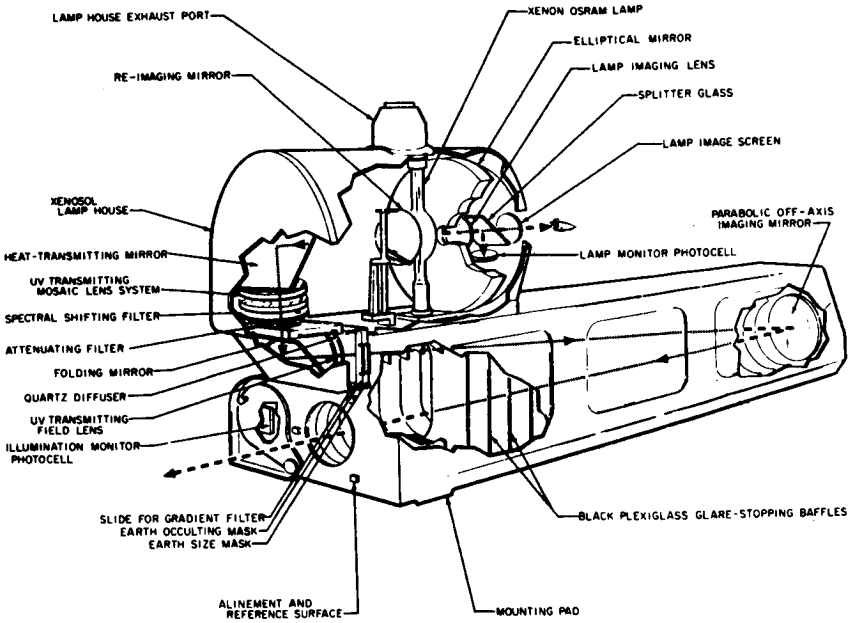


FIGURE 2.4—Planet-moon simulator.

distance. The simulator uses a mercury-xenon gas-discharge lamp as the source of light. The light from the lamp is reflected from a set of condenser mirrors and passes through a mosaic lens and a diffuser to provide uniform distribution of radiance across the planet image. The diffuse light then passes through masks which vary the size and shape of the simulated planet and through filters to adjust the spectral content of the light. The image of the planet formed by the mask is then imaged and collimated at infinity by an off-axis parabolic mirror. Accurate alinement is held throughout to prevent an angular error of the planet image. Intensity calibration is maintained by a lamp-monitor photocell and a photocell which may be swung into place in the position normally occupied by the instrument to be tested. The illumination and the size of the simulated planet-moon may be varied over a wide range by changing the current through the xenon lamp and adjusting the mask aperture without affecting the spectral content significantly.

A simulator for testing infrared horizon scanners has also been constructed (fig. 2.5). This simulator has an infrared target, which presents a heated surface surrounded by a background that may be cooled if desired. The heated target is split so that the planet terminator can be simulated, and the target image is collimated at infrared wavelengths by a 10-inch-diameter germanium lens. Since

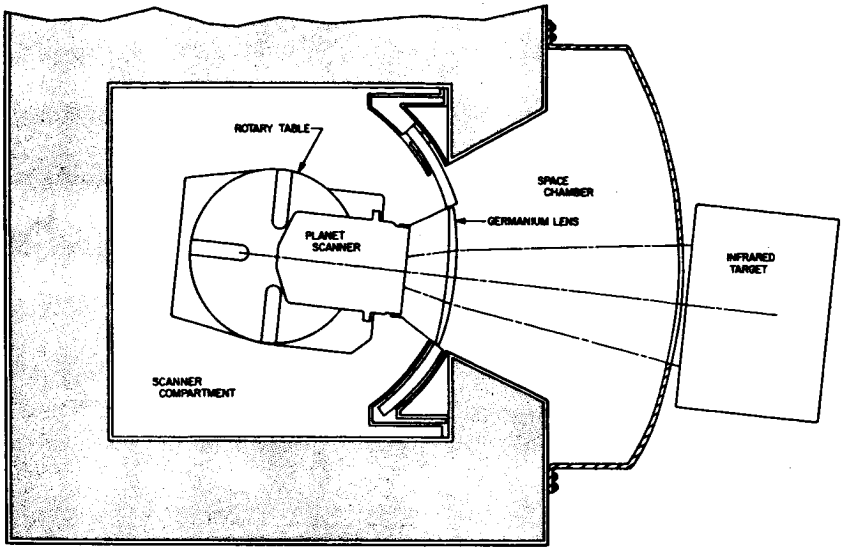


FIGURE 2.5—Infrared planet simulator.

it is not convenient to place the lens at the center of rotation of the unit under test, it may be shifted in position to maintain proper collimation during calibrations. The entire test turntable and horizon scanner is located in an environmental test chamber so that the characteristics of the scanner may be determined over wide variations in temperature.

A complete attitude-control system or an entire spacecraft may be tested in another facility called a celestial simulator (fig. 2.6). This simulator might be considered a space planetarium in that the Sun, Earth, and/or a planet may be simulated as they would be seen by a spacecraft thousands or millions of miles out in space. The simulator consists of a 40-foot-diameter hemispherical dome with a nonreflecting interior. A beam of collimated sunlight can be brought into the center of the dome by means of a 36-inch-diameter heliostat. The heliostat is unique in that the servodrive system derives its errors from a Sun sensor located in the bundle of light reflected into the simulator. During poor weather or at night, an artificial sun source using a 225–A carbon-arc light collimated with an off-axis parabolic mirror may be used. Although the artificial sun source can produce only about one-half the intensity of the Sun, it can simulate the spectrum seen above the Earth's atmosphere and can also vary the apparent diameter of the Sun to duplicate the characteristics seen on a planetary trajectory. Other optical and infrared targets can be positioned around the dome of the celestial simulator on a pair

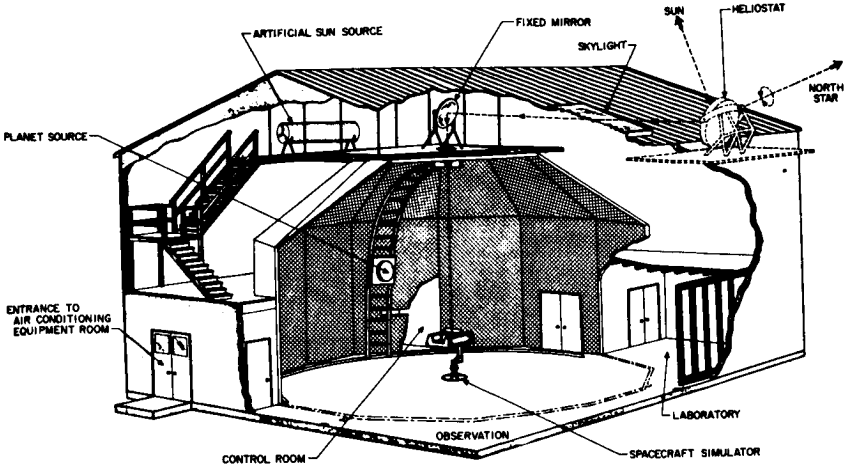


FIGURE 2.6—Celestial simulator laboratory.

of movable ladder cranes. The positions and rates of these secondary targets may be varied with two degrees of freedom. A simulated spacecraft supported on a 10-inch-diameter spherical gas bearing is located in the center of the dome. Sun sensors, Earth sensors, horizon scanners, or the entire attitude control system of a spacecraft can be tested in this facility.

CHAPTER 3

Gyroscopes

A NECESSARY PART of most guidance systems is the gyroscope (refs. 3–7). Gyros are used to provide an angular reference for attitude and rate information in a guidance system as well as a method for measuring acceleration, which will be shown in chapter 4. The accuracy of an inertial guidance system is ultimately limited by the inherent errors and sensitivity of the inertial measuring instruments.

SIMPLE TWO-DEGREE-OF-FREEDOM GYROS

The basis of all gyros and accelerometers is, in one way or another, Newton's laws of motion. A body containing significant mass, inertia, or momentum is established in a given orientation, and the motion of the missile or spacecraft is measured relative to the body which retains its position or orientation in inertial space. The simplest form of a gyro is shown in figure 3.1. It consists of a rapidly spinning wheel or rotor suspended by some sort of mounting which allows it to have two degrees of freedom relative to the base on which it is supported. A common form of this mounting is two gimbal rings forming a suspension similar to a universal joint. For the purpose of discussion, it is assumed that the gimbal bearings of the mount and the pickoff have no friction, and that all the parts are perfectly balanced.

Once the rotor is set spinning with the spin axis in an arbitrary direction, it will remain in that direction even though the outer case of the device is rotated or translated relative to the spinning rotor. The spinning rotor will remain in its initial orientation in inertial space unless some external torque is applied to the rotor perpendicular to the spin axis. This torque will cause the rotor to rotate about a third axis in quadrature, or as it is more commonly known, the rotor will precess.

When real hardware is used, the rotor and gimbal bearings have friction, the rotor and gimbals are unbalanced, the pickoff has friction and/or compliance, and all these characteristics change with time, temperature, and other environmental factors. In a practical two-degree-of-freedom gyro similar to the one shown, the torques re-

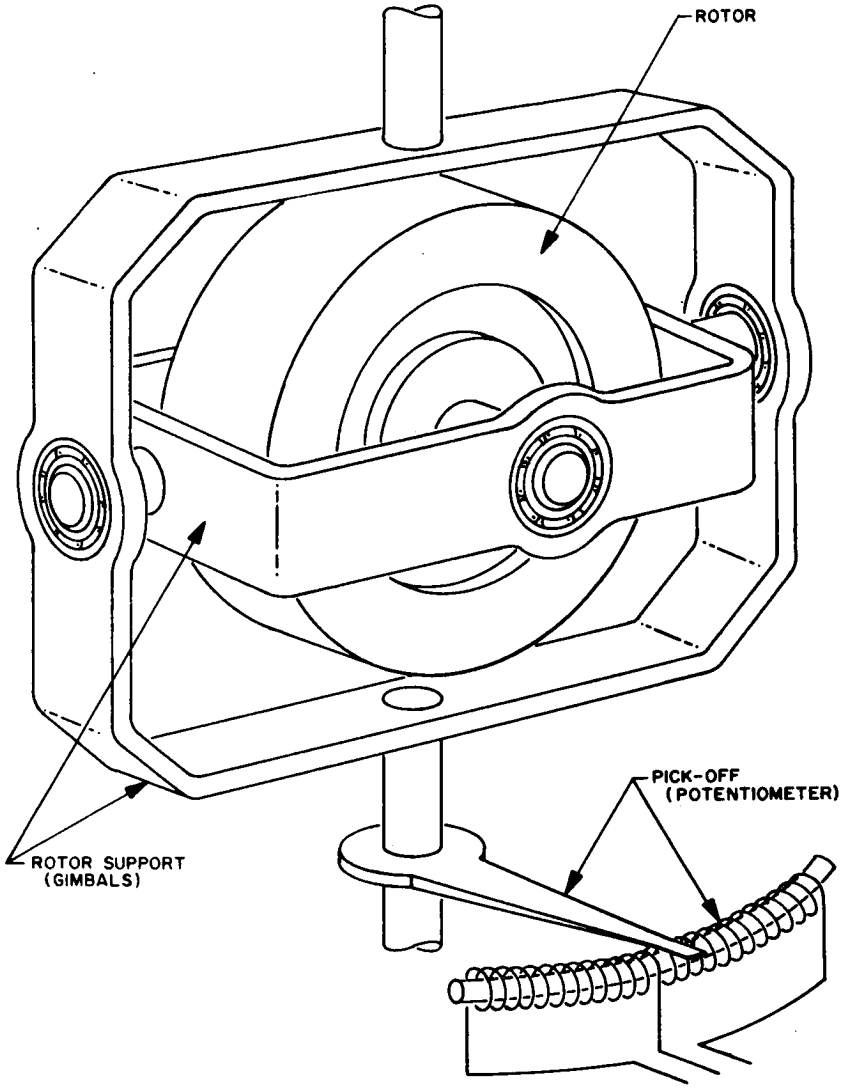


FIGURE 3.1—Principal parts of a gyroscope.

sulting from these undesirable error sources produce a precession of the spin axis of about 0.25 to 1.0 deg/min. This unwanted spin-axis precession is generally known as gyro drift.

This type of gyroscope is commonly used as a directional gyro or vertical gyro for aircraft or as an attitude gyro for simple missile systems. In either case, it can be used to provide an electrical output

used by the autopilot system, or in the case of manned aircraft or spacecraft, to provide a visual attitude reference such as a gyrocompass or an artificial horizon. The accuracy limitations of this type of mounting or suspension system usually restrict this type of gyro to a relatively simple, inaccurate orientation reference. This reference is normally required to be reset from time to time by some external source such as magnetic north, gravity vertical, or a radio direction. There are some extremely precise two-degree-of-freedom gyros using very low torque suspension systems which can be used for more accurate purposes. These gyros will be discussed at the end of this chapter.

SINGLE-DEGREE-OF-FREEDOM GYROS

A method for reducing the stray torques causing gyro drift was developed during the late 1940's and early 1950's using a single-degree-of-freedom gyro. A brief review of the basic laws of gyroscopic motion is in order before describing this type of gyro.

The fundamental relationship describing the motion of a rotating wheel under the influence of an external torque perpendicular to the axis of spin of the body is

$$\mathbf{T}=\boldsymbol{\omega}\times I_s\boldsymbol{\omega}' \tag{3.1}$$

where

$\mathbf{T}$	torque about the input axis
I_s	moment of inertia about the spin axis
$\boldsymbol{\omega}'$	angular velocity about the spin axis
$\boldsymbol{\omega}$	angular velocity about the precession axis

The angular momentum of the wheel is represented by

$$\mathbf{H}=I_s\boldsymbol{\omega}' \tag{3.2}$$

so that

$$\mathbf{T}=\boldsymbol{\omega}\times\mathbf{H} \tag{3.3}$$

The relationship between the angular momentum vector, the torque vector, and the precession vector may be seen in figure 3.2.

Similarly an angular rate about the T-axis would produce a torque about the precession axis. To avoid confusion about whether the input to the gyro is a torque or an angular rate, the axes of a single-degree-of-freedom gyro are usually labeled the spin axis or spin-reference axis (SRA); the input axis (IA); and the precession or output axis (OA). These axes are shown on figures 3.3 and 3.4.

In the single-degree-of-freedom gyro shown in figure 3.1, the spinning wheel with its set of spin bearings has only one additional degree of freedom with respect to the gyro case. An angular rate ω_1 about

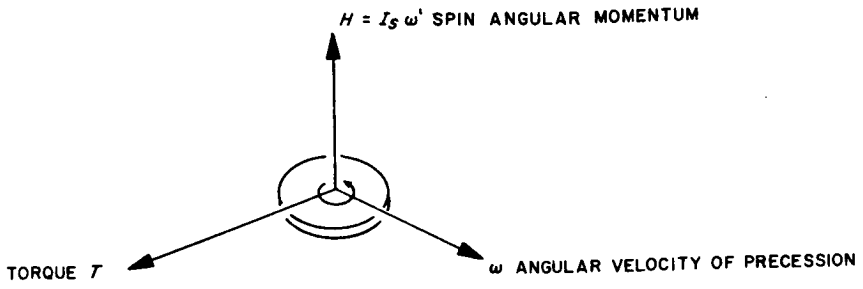


FIGURE 3.2—Relationship among angular momentum vector, torque vector, and precession vector.

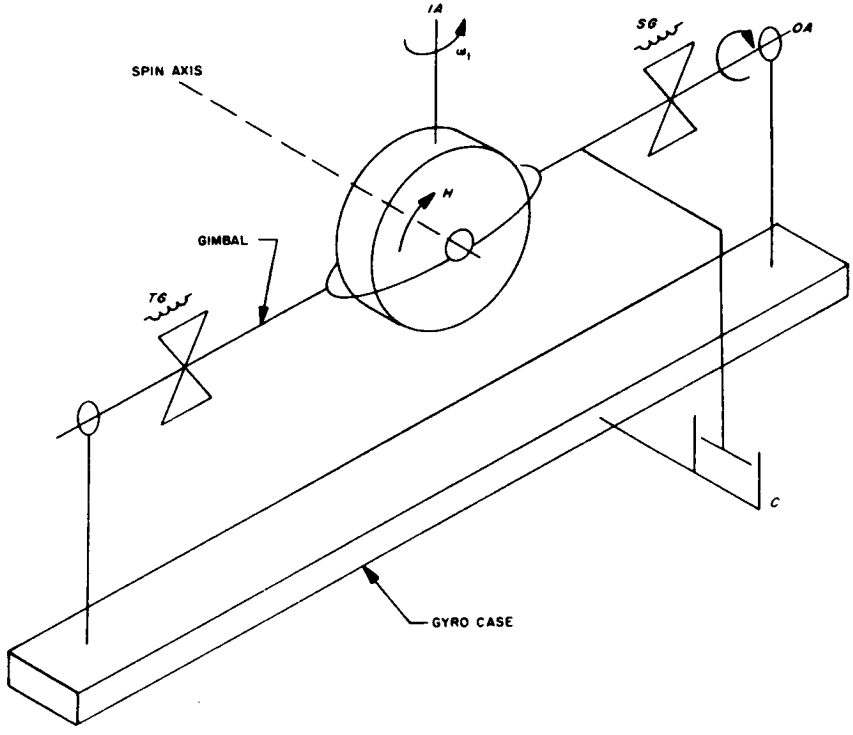


FIGURE 3.3—Single-degree-of-freedom gyro.

the input axis (IA) will cause a precession torque about the output axis (OA). The torques opposing any gyroscopic torque about the output axis are due to the inertia, viscous damping, and spring-reaction torques acting on this axis. Thus, the sum of all the torques acting on the OA is

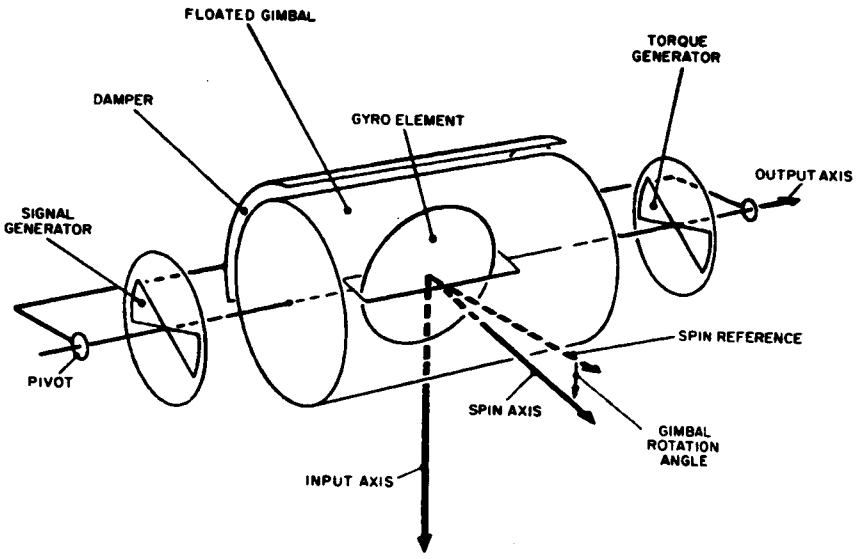


FIGURE 3.4—Single-degree-of-freedom integrating gyro.

$$T = H\omega_1 = I_0\ddot{\theta} + C\dot{\theta} + K\theta \tag{3.4}$$

where

ω_1	rate about the IA
I_0	inertia about the OA
C	damping about the OA
K	spring constant about the OA
θ	angular precession or rotation about the OA

If the spring constant K is made large compared with the inertia and damping, the gyro has the following characteristics:

$$H\omega_1 \approx K\theta \tag{3.5}$$

thus

$$\theta \approx \frac{H}{K}\omega_1 \tag{3.6}$$

or, the output angle is directly proportional to the input rate. Thus, the gyro becomes a rate-measuring instrument. In this configuration, it is known as a spring-restrained rate gyro. Equation (3.6) is exact for low frequencies, but is in error near and above the resonant frequency determined by I_0 and K.

Some form of angular position pickoff, shown as signal generator (SG), may be employed to provide an electrical output. A direct visual output in the form of a pointer may also be used, as in the common turn-and-bank indicator employed in aircraft. This type

of gyro is commonly used to provide a rate or damping signal to stabilize an autopilot system. For this type of gyro, the common range of input rates varies from a few degrees per minute to hundreds of degrees per second, although in any one instrument the linearity and null errors of the gyro would probably be from 1 to 5 percent of full-scale rate. The typical error sources of this type of gyro are pickoff and spring nulls not exactly alined, unbalance of the rotor or gimbal, and damping or other highly temperature-dependent characteristics.

If the damping constant is made large compared to the inertia and spring constant, the gyro has entirely different characteristics

$$H\omega_1 \approx C\dot{\theta} \tag{3.7}$$

thus

$$\dot{\theta} \approx \frac{H}{C}\omega_1 \tag{3.8}$$

or integrating

$$\theta \approx \frac{H}{C}\phi_1 \tag{3.9}$$

where ϕ_1 = the angle change about the IA.

If the ratio H/C is made unity, as it is in many gyros of this type, the angular rotation of the gyro about the output axis is equal to the angular rotation about the input axis. The gyro thus becomes a form of position gyro known as a single-degree-of-freedom integrating gyro. Another more commonly used term for this type of gyro is a hermetic integrating gyro (HIG), since the wheel assembly is hermetically sealed in a cylindrical or spherical canister and the damping is provided by viscous shear of a fluid in which it is immersed.

A further advantage of this type of construction is that the canister, or float as it is usually called, is made neutrally buoyant in the damping fluid, so that the bearings on the ends of the output axis do not have to support any weight, and thus can be of a very low friction type. A common type of OA bearing in a HIG gyro is a pivot and sapphire jewel such as that used for the low-friction bearings of the balance wheel in a watch. Thus, both the coefficient of friction and the supported mass are very low (zero except for errors in buoyancy). The resulting friction is so low that if a sled had the same coefficient of friction compared to its total weight as does a HIG gimbal float, it could coast down an incline having a drop of only 1 ft/1000 miles.

Other advantages which stem from this type of gyro construction is that the damping fluid also provides a relatively uniform temperature environment and damps out structural vibration of parts immersed in the fluid caused by the propulsion system or aerodynamic

forces. HIG gyros are normally temperature controlled to 1° F, or better. This provides a uniform, known environment for control of unbalance due to differences in the coefficient of expansion of materials, and also because of the density of the flotation (damping) fluid (and thus the buoyant force) change with temperature.

Gyros of the HIG type are made in a variety of sizes from less than 1 inch in diameter to larger than 5 inches. The better units have less than 1 deg/day drift when used in a system where known constant-error terms are compensated for. This type of gyro is used in most of the missiles and spacecraft where accurate attitude information is required.

OTHER GYRO TYPES

Many other forms of gyroscopes have been suggested over the years, many of which have been put into practice. Essentially, any way that momentum can be stored and its orientation read out will form a type of gyro.

Two other types of gyros used in navigation, guidance, and control systems today deserve special mention. These are the two-degree-of-freedom, free-rotor gyro, and the two-degree-of-freedom floated gyro.

The two-degree-of-freedom, free-rotor gyro consists of a rotating inertial mass supported by some form of a spherical low-friction bearing. Some rotors take the form of a flanged wheel with a section of a sphere at its hub. Others take the form of spheres which may or may not be hollowed out to provide a preferred moment of inertia axis. The forms of support bearings vary from hydrodynamic and hydrostatic gas bearings to exotic types employing electrostatic or electromagnetic fields to support the rotating mass. The flanged-wheel configuration provides two degrees of freedom over only a small limited range, while the spherical-rotor type has unlimited angular freedom. Accuracy of this class of gyro is limited by similar types of errors present in the HIG gyro, namely, unbalances and stray torques introduced by imperfections in the low-friction support system. A few recent examples of gyros of this type have proven to be more accurate than the HIG type, although requiring a greater amount of complexity in the associated equipment.

The two-degree-of-freedom floated gyro is fundamentally similar to the single-degree-of-freedom type, except that the float is normally spherical. A gimbal ring provides the second degree of freedom relative to the case. Significant differences between this type of gyro and the HIG are that the flotation fluid is not used to provide viscous-shear integration, and that fine torsion wires are normally used for gimbal support instead of pivots and jewels. Because of the

finite spring constant of the torsion wires, there is only limited angular freedom between the inner float and the outside case in order to reduce the effects of their torque on the orientation of the spinning mass. Gyros of this type have been built having accuracy comparable to that of the HIG configuration.

Many other existing forms of gyroscopes include vibrating tuning forks, torsional pendulums, rotating bodies of liquids and gases, nuclear gyros employing the angular momentum of nuclei of certain atoms, and laser gyros using the phase difference of two beams of coherent light. However, no matter how sophisticated the device is, it is still subject to friction and small extraneous torques which will introduce significant errors.

CHAPTER 4

Accelerometers

AN ACCELEROMETER (refs. 3–7), as its name implies, is a device to measure acceleration. An accelerometer is normally employed to sense translational motion, as contrasted with a gyro which senses rotational motion. As in the case of the gyro, the principles on which the accelerometer is based are Newton's laws. The basic form of an accelerometer is some seismic mass on which a force must act to make the mass keep up with the accelerating vehicle. The function of the accelerometer is to measure this force, which, obeying the law $F=MA$, is a measure of the applied acceleration.

The simplest form of accelerometer is illustrated in figure 4.1. It consists of a seismic mass, a frictionless slide bearing, and a spring restraint which obeys Hooke's law. The deflection of the spring, being proportional to the force, becomes a direct measure of the acceleration acting on the case. Although there is a wide variety of acceleration-sensing devices, the feature common to all is that they measure the force between the outer case of the instrument and a

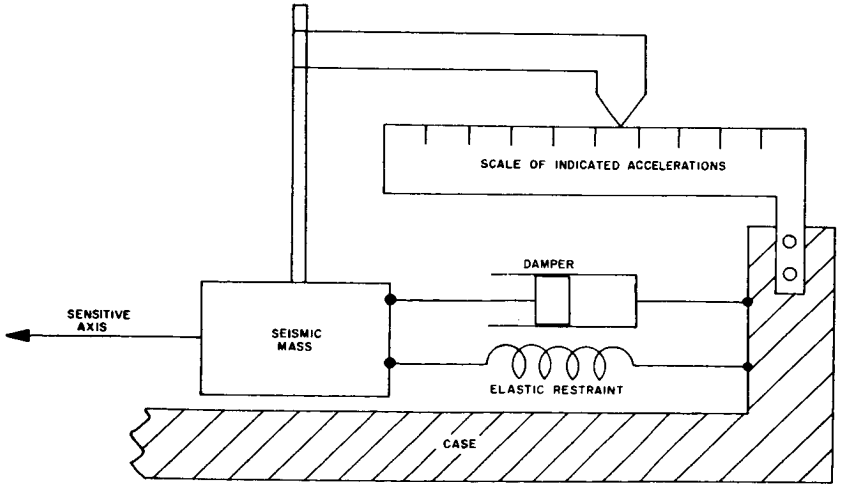


FIGURE 4.1—Simple form of accelerometer.

238-635 O—67——3

seismic mass. The main limiting factors on the accuracy of an accelerometer are the extraneous forces caused by the support system and pickoff compared with the acceleration-induced forces acting on the seismic mass. The type of accelerometer shown in figure 4.1 is actually used for some low-accuracy purposes such as telemetry, but it is not suitable for navigation or guidance purposes. Telemetry or instrumentation accelerometers quite often combine the spring-constraint and the force- or deflection-measuring sensor by using unbonded strain-gage wires or piezoelectric crystals for supporting the seismic mass.

FORCE-BALANCE ACCELEROMETERS

A force-balance accelerometer is an instrument in which the acceleration force acting on the seismic mass is counteracted by an equal and opposite force without requiring the mass to cause an actual deflection of the measuring device, as in the spring-mass accelerometer described above. An example of a force-balance accelerometer is shown in figure 4.2. The acceleration force acting on the mass is sensed by an electrical pickoff. The output voltage from this pickoff is amplified by a high-gain amplifier, and the output of this amplifier is connected to a force coil. The force of this coil is directly proportional to the current flowing through it, so a precision series resistor is added to the circuit to measure this current.

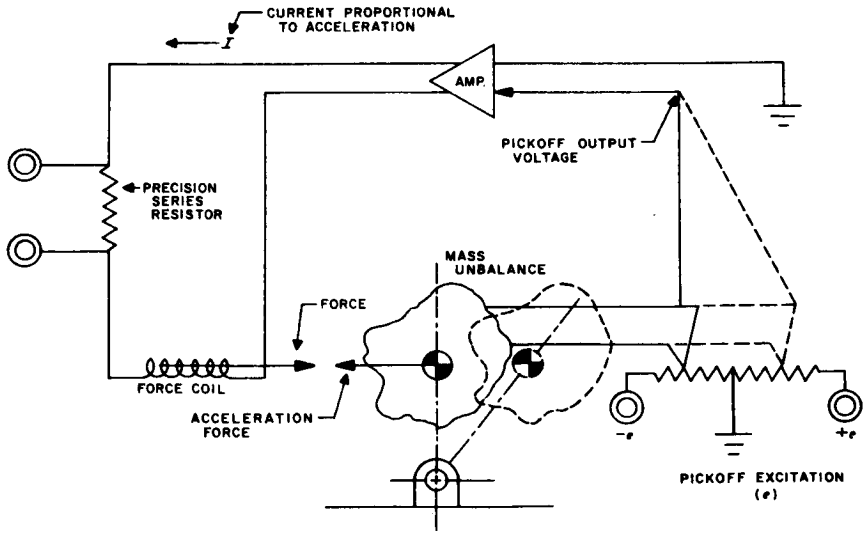


FIGURE 4.2—Accelerometer model.

Any error between the acceleration-induced force and the force coil causes a motion of the pickoff, resulting in a correction of the torque balance. Thus, the voltage drop across the precision resistor becomes a measure of the acceleration. In practice, the gain of the amplifier and the sensitivity of the pickoff can be made very high. Thus, deflections in the vicinity of a few millionths of an inch will produce full output torque. This type of an accelerometer is electrically equivalent to a very stiff spring. However, the only linear or precision device required is the force coil and series resistor. As in other inertial sensing instruments, however, stray forces acting on the seismic mass cannot be distinguished from those induced by acceleration. The effects of pickoff and amplifier nonlinearity and friction are greatly reduced by using high amplifier gain.

One form of the force-balanced accelerometer uses a very weak spring as a pivot for supporting a pendulous mass containing a torque coil centered in a linear magnetic field provided by permanent magnets. Another form of the force-balance pendulum is shown in figure 4.3. The unit is very similar to the HIG gyro, except that the wheel is replaced by a pendulous mass. The float is still made neutrally buoyant and is supported by pivots and jewels. A signal-generator pickoff is located at one end of the float and a torque generator at the other. A cutaway drawing of an actual force-balance pendulum accelerometer is shown in figure 4.4.

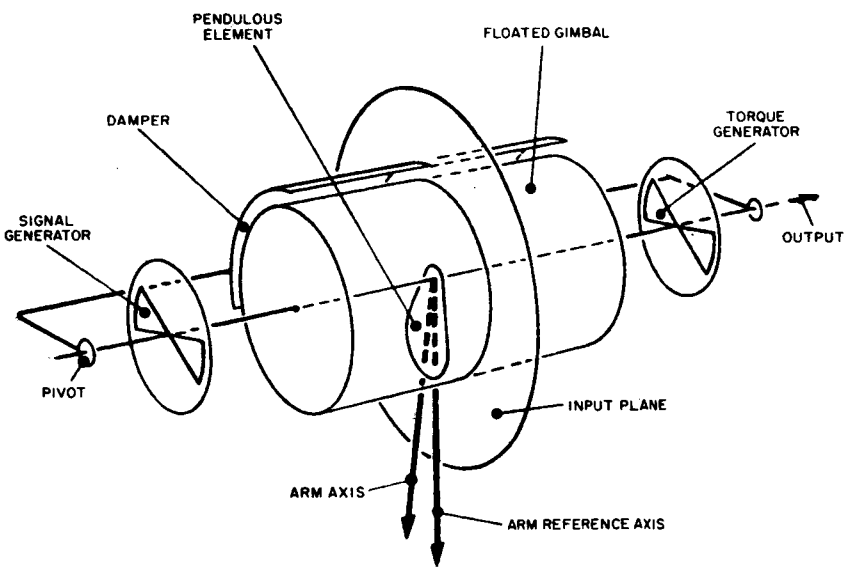


FIGURE 4.3—Single-degree-of-freedom pendulum unit.

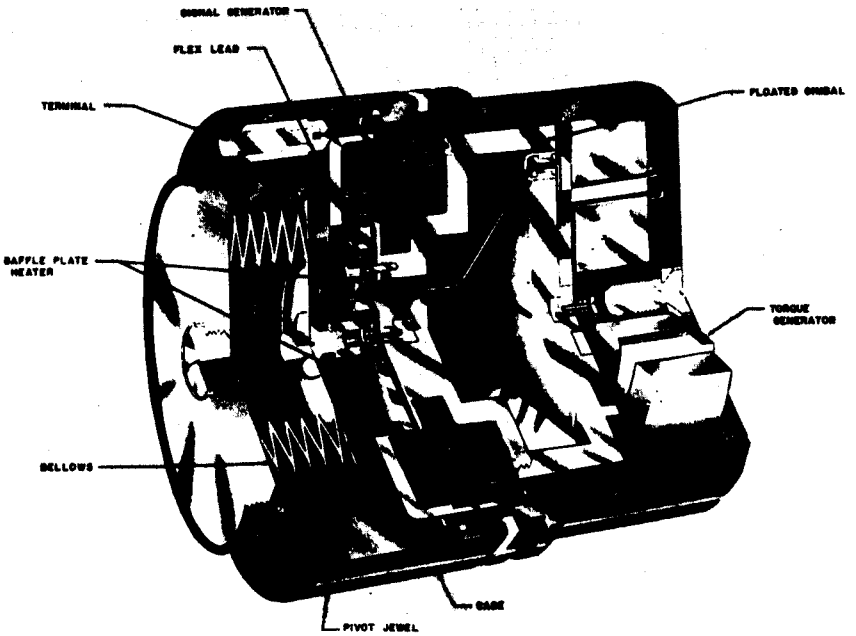


FIGURE 4.4—Force-balance pendulum accelerometer.

This same accelerometer may be used in a digital or pulse system by changing the electronics used with it. If every time the position error of the seismic mass builds up beyond a certain predetermined level, a pulse of uniform height and length is fed into the torquer; then the average pulse rate is proportional to the acceleration. The total number of the pulses is proportional to the time integral of the acceleration, or it is equal to the velocity acting on the body. This can be of some advantage in system applications, as will be shown later. This version of accelerometer is known as a Pulse Integrating Pendulous Accelerometer (PIPA).

GYRO-TYPE ACCELEROMETERS

The fundamental gyro equation, $\mathbf{T}=\boldsymbol{\omega}\times\mathbf{H}$, is the basis for the gyro-type accelerometer. In figure 4.5, a spinning gyro wheel with angular momentum H is shown mounted off its center of gravity, thus producing a torque around the inner gimbal pivot axis. This torque, T, will cause an angular rotation at precession rate ω about the vertical outer-gimbal axis. Since

$$T=MlA \tag{4.1}$$

where

M mass of the wheel and inner gimbal
l distance from inner gimbal pivot to the center of gravity
A acceleration acting along the vertical axis

and

$$T=H\omega \tag{4.2}$$

then

$$\omega=\frac{Ml}{H}A \tag{4.3}$$

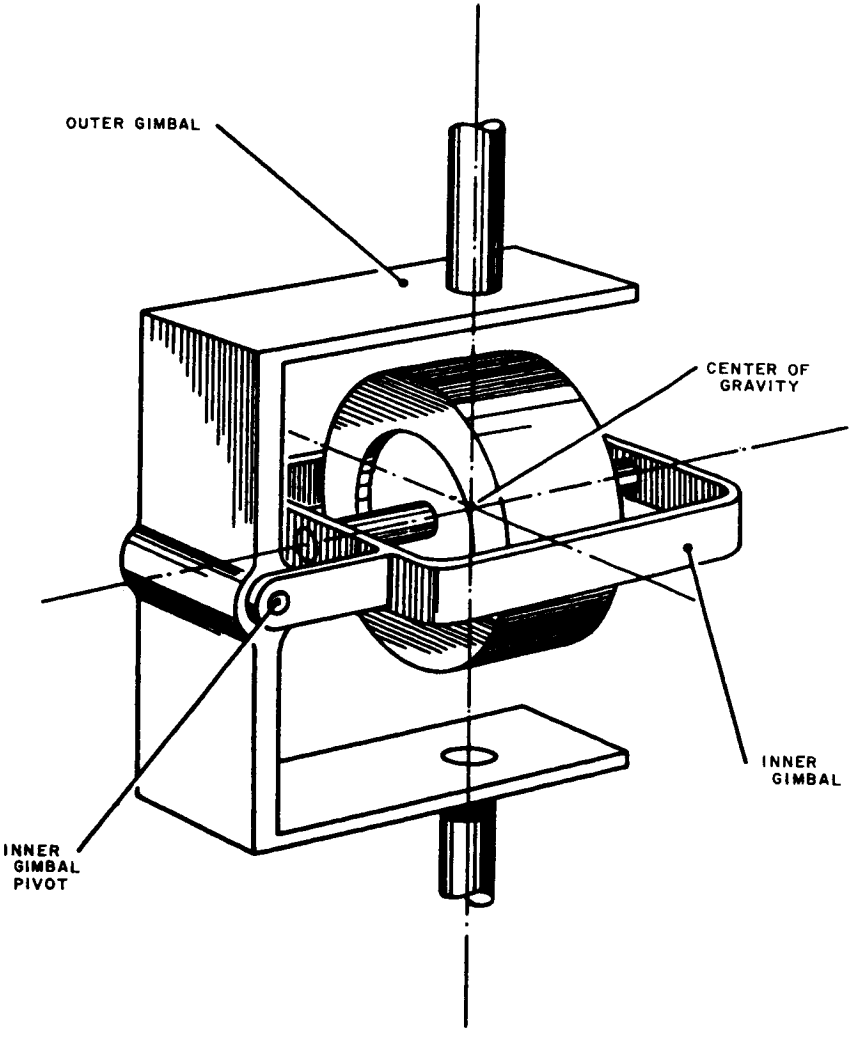


FIGURE 4.5—Pendulous gyro.

Thus, the precession rate about the vertical axis is directly proportional to the acceleration acting along that axis. The integral of the precession rate, or the number of revolutions made by the outer gimbal, is proportional to the integral of the acceleration, or the velocity along the vertical axis. Thus, this type of accelerometer is also a form of integrating accelerometer. As in the simple spring-mass accelerometer, the energy required to overcome friction and accelerate the outer gimbal must come directly from the inertial element, so this form of the gyro accelerometer is subject to errors.

A more refined version of this type of gyro accelerometer is shown in figures 4.6 and 4.7. The gyroscope (fig. 4.6) consists of a HIG, modified to include a pendulous mass mounted on its spin axis. The pendulous gyro is mounted on a motor-driven turntable, as shown in figure 4.7.

An acceleration force acting on the unbalanced mass in the gyro causes a torque about the OA. An error is sensed by the signal generator which is amplified to provide a large current to the turntable drive motor. The rotation rate of the turntable about the IA of the gyro causes a precession torque about its OA. The entire system is in equilibrium when the torque caused by acceleration acting on the pendulous mass is equal and opposite to the precession torque generated by the turntable rotation as in equation (4.3). Thus, the turntable speed is proportional to acceleration and the total revolutions of

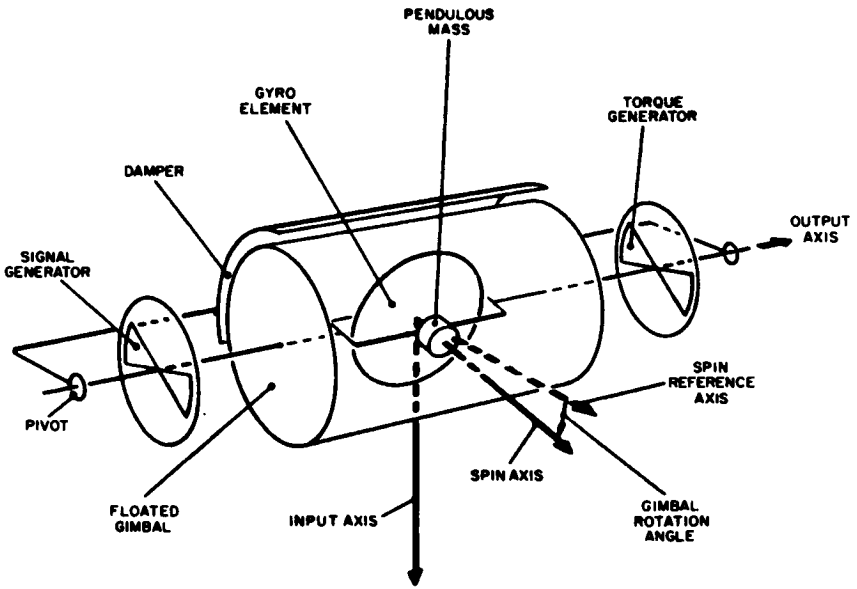


FIGURE 4.6—Single-degree-of-freedom pendulous-integrating gyro.

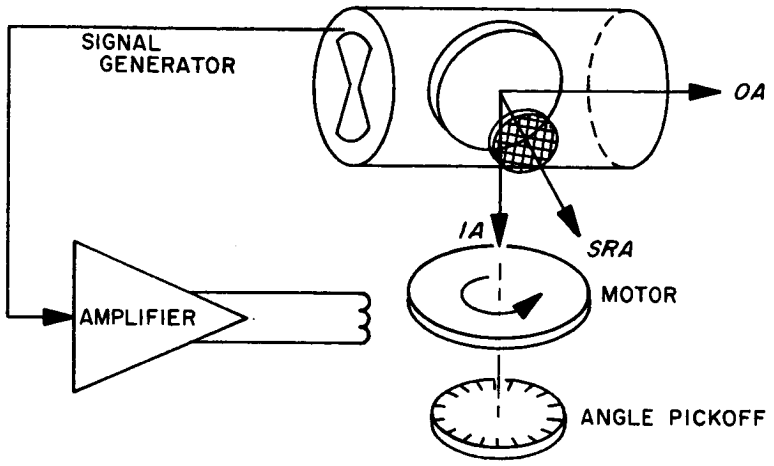


FIGURE 4.7—Pendulous gyro mounted on motor-driven turntable.

the turntable are proportioned to velocity along the IA of the gyro. In this case, however, the energy required to overcome friction and inertia of the turntable is provided by its motor, and if the amplifier gain is high, insignificant errors will result from this cause. This type of accelerometer is known as a Pendulous Integrating Gyro Accelerometer (PIGA).

VIBRATING-STRING ACCELEROMETER

In the form of accelerometer shown in figure 4.8, a seismic mass is suspended between two strings which are under longitudinal tension and are vibrating transversely. Longitudinal acceleration forces cause an increase in the tension of one string and a corresponding decrease in the other. The frequency of vibration of a string in tension is

$$f^2=\frac{T_L}{4\sigma L^2} \tag{4.4}$$

where

T_L longitudinal tension in the string
σ density of the string per unit length
L length of the string

The frequencies of vibration of the strings are

$$f_1^2=\frac{T_{L1}}{4\sigma L^2} \qquad f_2^2=\frac{T_{L2}}{4\sigma L^2} \tag{4.5}$$

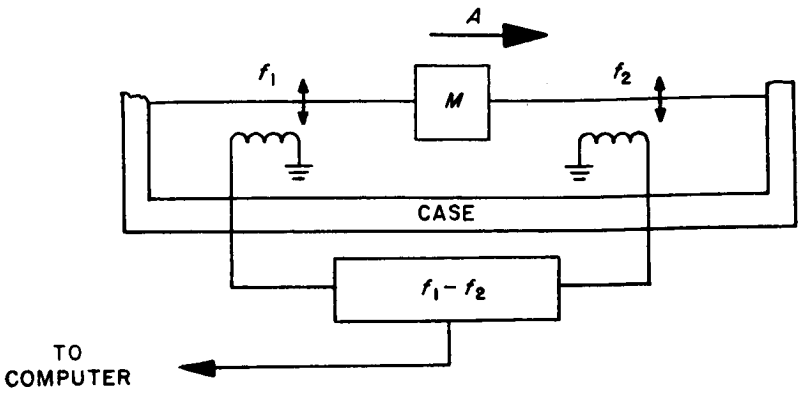


FIGURE 4.8—Vibrating-string accelerometer.

The difference in tension of the two strings, $T_{L1} - T_{L2}$, is the acceleration force acting on the seismic mass and thus is equal to MA. Thus:

$$f_1^2 - f_2^2 = \frac{T_{L1} - T_{L2}}{4\sigma L^2} = \frac{MA}{4\sigma L^2} \tag{4.6}$$

Factoring

$$(f_1 - f_2)(f_1 + f_2) = \frac{MA}{4\sigma L^2} \tag{4.7}$$

Thus

$$f_1 - f_2 = \frac{MA}{(f_1 + f_2)4\sigma L^2} \tag{4.8}$$

Thus, the difference frequency of the two strings is directly proportional to the applied acceleration if the sum frequency is constant. Because of temperature changes, the term $(f_1 + f_2)$ does not remain constant, and the effect of this inaccuracy must be removed by a computer or through control of the string tension by some means. Some higher order nonlinearities are introduced into this type of accelerometer by coupling between the strings, by string termination effects, and by the transverse support required to maintain the mass centered between the string terminations under the effects of transverse vibration or acceleration.

This type of accelerometer is an almost ideal type of inertial sensor in that it provides a direct translation from the acceleration being sensed to a frequency. Counting the total number of cycles gives a direct digital measure of vehicle velocity without the need to convert from an analog to a digital quantity. Some of the mechanical difficulties mentioned above tend to counteract the benefits, making this accelerometer system about as difficult to manufacture as the PIPA or PIGA.

OTHER TYPES OF ACCELEROMETERS

A large variety of accelerometers has been conceived and built. Generally, accelerometers have been developed around the means by which a force can be generated and measured, particularly forces which are in turn a function of velocity (such as viscous shear) or acceleration (inertial-reaction force). The advantage of the viscous or inertial-reaction techniques is that they provide one or two integrations in the instrument itself, thereby providing a direct measure of velocity or distance without using a computer.

The eddy-current velocimeter is shown in figure 4.9. It consists of a pendulous mass connected to an eddy-current drag cup supported on a low-friction bearing. Acceleration forces cause a rotation of this mass which is sensed by a pickoff. This signal is amplified and used to drive a motor attached to the eddy-current generator. The induced eddy currents produce a torque on the drag cup proportional to the motor speed. The system reaches equilibrium when the eddy-current torque equals the acceleration-induced torque. As in the other integrating accelerometers, the number of revolutions of the motor is then proportional to the velocity of the vehicle. A similar type of instrument could be made using the viscous shear of a fluid coupling to take the place of the eddy-current device.

Double integration may be obtained in a similar type of arrangement if both the motor rotor and stator are supported on bearings. Linear acceleration acting on an unbalanced mass on the stator housing would introduce an error used to drive the motor. The angular acceleration imparted to the rotor would create an equal and opposite reaction torque to the motor stator to cancel out the effects of ex-

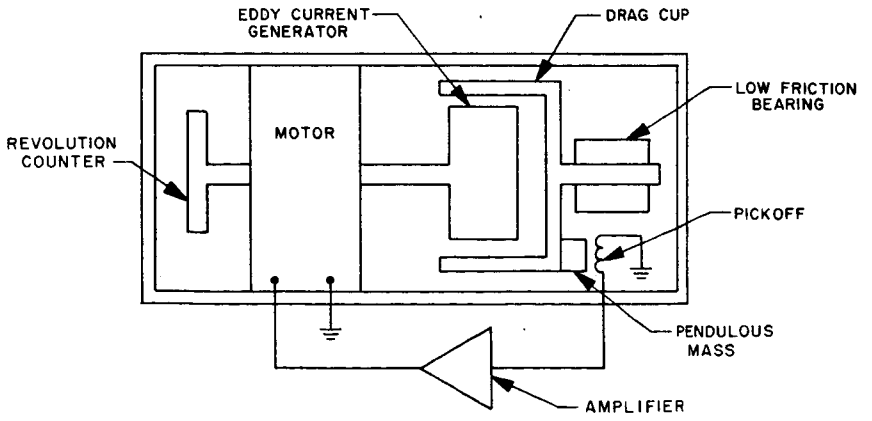


FIGURE 4.9—Eddy current velocimeter.

ternally applied acceleration to the pendulous mass. Thus, the rotor acceleration is proportional to the external linear acceleration, and so, integrating twice, the rotor angle is proportional to the distance the vehicle has traveled.

Other forms of accelerometers too numerous to mention have been built using such techniques as hydraulic pressure, electrostatic forces, masses freely suspended in viscous fluids, bubbles in fluids, and quartz fibers. However, the ones described in detail here are the most common types in use. One conclusion which might be drawn is that there is a wide selection of equally satisfactory ways to measure acceleration and relatively few ways to measure angular changes by a gyro.

CHAPTER 5

Servomechanisms

A SERVOMECHANISM (ref. 8) is a form of control system in which a comparison is made between the desired output and the actual output, and the error between these two quantities is used to provide the control. Common terms used to describe this form of system are "closed loop" or "feedback" control.

CONTROLLERS

Many of the forms of control used during our daily tasks are examples of open-loop control. In this type of control, the output is controlled directly in some relationship to the input. Typical examples are the volume control on a radio or the throttle of an automobile. The input signal to the control would be the setting of the control to the desired position, and the controlled function, such as the volume or the amount of gasoline being fed to the engine, is a function of the setting of the control.

The advantage of this type of open-loop control is that it is very simple and usually fairly reliable. The disadvantage is that the controlled output normally is also a function of many other things as well, such as the strength of the radio station, or, in the case of the throttle, the slope of the road and the setting of the gears of the transmission. Thus, the open-loop controller is subject to errors from variations of the intermediate components, load changes on the output, or variations in the input in the device being controlled.

CLOSED-LOOP SYSTEMS

To eliminate the disadvantages of this simple type of control, it is often necessary to use a form of closed-loop control. The open-loop system is often turned into a version of a feedback control system by including a human operator as the sensing, feedback, and differencing part of the loop. Thus, the human operator can sense the speed of the automobile by looking at the speedometer, and make an adjustment in the throttle position until the desired speed is obtained. A block

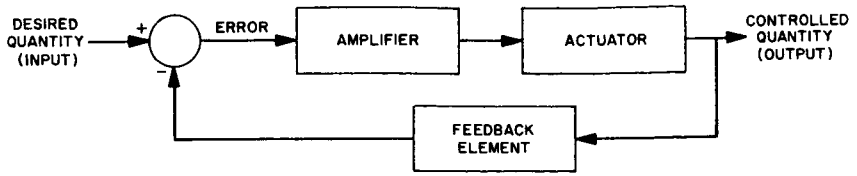


FIGURE 5.1—Block diagram of basic servomechanism.

diagram describing a basic servomechanism or closed-loop control system is shown in figure 5.1. This block diagram may be simplified for analysis purposes to that shown in figure 5.2.

The transfer function of the servomechanism is the ratio θ_o/θ_i, where θ_o is the error, ϵ, multiplied by the gain, A.

$$\theta_o = A\epsilon \tag{5.1}$$

The error, ϵ, is composed of the difference of two quantities, θ_i, and the product of the output, θ_o, and the feedback gain, β.

$$\epsilon = \theta_i - \beta\theta_o \tag{5.2}$$

Combining equations (5.1) and (5.2)

$$\theta_o = A(\theta_i - \beta\theta_o)$$
$$\theta_o = A\theta_i - A\beta\theta_o$$
$$\theta_o + A\beta\theta_o = A\theta_i$$
$$\theta_o(1 + A\beta) = A\theta_i$$

thus

$$\frac{\theta_o}{\theta_i} = \frac{A}{1 + A\beta} \tag{5.3}$$

It may be seen by examining equation (5.3) that the output may be made equal to the input (or the transfer function θ_o/θ_i equals unity)

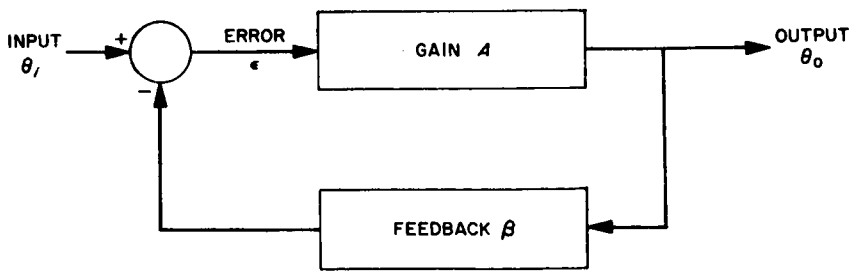


FIGURE 5.2—Simplified closed-loop control system.

if the gain of the loop, A, is very large and the feedback gain, β, is 1. It should be noted also that differences in gain of the forward part of the loop do not affect the transfer function if the gain is sufficiently high. As an example, if

$$A=1000$$

$$\beta=1$$

$$\frac{\theta_o}{\theta_i}=\frac{1000}{1001}=0.9990$$

Assuming a gain change of the amplifier of 10 percent

$$A=1100$$

$$\beta=1$$

$$\frac{\theta_o}{\theta_i}=\frac{1100}{1101}=0.9991$$

or a change of one part in the fourth significant place. Thus, the transfer function (or the accuracy of control) may be seen to be quite independent of the gain (or linearity or other characteristics) as long as the loop gain is sufficiently high. In some types of precision servomechanisms, such as analog computers to be described later, the loop gain may be as high as 10^6.

Examination of equation (5.3) reveals other characteristics of a servomechanism. If the feedback gain, β, changes, then the transfer function is changed in approximately the same ratio. Thus, the nature and stability of the feedback element essentially determines the characteristics of the entire servomechanism. It may also be seen that if the quantity $A\beta$ has the value of -1, the denominator of the equation goes to zero and the equation becomes indeterminate. In actuality, the closed-loop system becomes unstable when the feedback is positive ($A\beta \rightarrow -1$) and the system will be divergent with time. Thus, it can be seen that in contrast to the open-loop controller, the servomechanism can be made very accurate and independent of changes in the forward loop but with a sacrifice of stability.

DYNAMIC CHARACTERISTICS

An elementary closed-loop position servo is shown in figure 5.3. The servo loop consists of a potentiometer positioned by input knob, θ_i, providing an input signal to the difference junction. The error signal, e_1, is amplified by a gain of A and provides an input voltage e_2 to the motor. The motor drives an inertia load attached to output potentiometer, θ_o, generating feedback voltage, e_o.

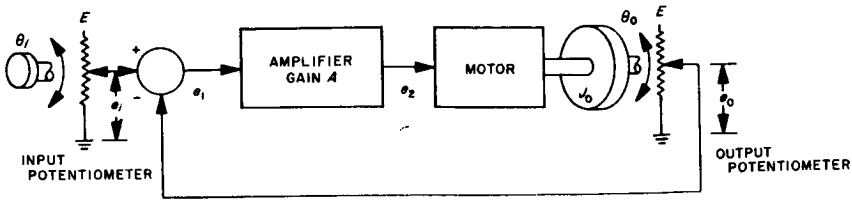


FIGURE 5.3—Elementary closed-loop position servo.

The input potentiometer converts the input angle to a voltage by the following relationship

$$e_i = K_1\theta_i \tag{5.4}$$

Similarly, the output potentiometer has these characteristics

$$e_o = K_1\theta_o \tag{5.5}$$

where, in V/rad,

$$K_1 = \frac{E}{\theta_{\max}} \tag{5.6}$$

The output of the difference junction, e_1, is the difference between the input and the output signals.

$$e_1 = e_i - e_o = K_1(\theta_i - \theta_o) \tag{5.7}$$

This difference is amplified by factor A and appears at the terminals of the motor.

$$e_2 = Ae_1 = AK_1(\theta_i - \theta_o) \tag{5.8}$$

The motor-torque characteristics may be described by

$$T = K_2e_2 - K_3\frac{d\theta_o}{dt} \tag{5.9}$$

where T is the torque output of the motor; the first term of the equation, K_2e_2, represents the torque as a function of applied voltage; and the second term, $K_3(d\theta_o/dt)$, represents the torque/speed relationship.

The torque is applied to an inertia load J_o. The summation of torques is then

$$K_2e_2 - K_3\frac{d\theta_o}{dt} = T = J_o\frac{d^2\theta_o}{dt^2} \tag{5.10}$$

Substituting equation (5.8) for e_2,

$$K_2K_1A(\theta_i - \theta_o) - K_3\frac{d\theta_o}{dt} = J_o\frac{d^2\theta_o}{dt^2} \tag{5.11}$$

Combining and simplifying

$$\frac{d^2\theta_o}{dt^2}+\frac{K_3}{J_o}\frac{d\theta_o}{dt}+\frac{K_1K_2A}{J_o}\theta_o=\frac{K_1K_2A}{J_o}\theta_i \tag{5.12}$$

This equation has the form of a second-order linear differential equation with constant coefficients. This equation appears in similar form in almost all servo problems, so an understanding of its properties will be valuable. A generalized version of this equation is as follows:

$$\frac{d^2\theta}{dt^2}+2\zeta\omega_\eta\frac{d\theta}{dt}+\omega_\eta^2\theta=f(t) \tag{5.13}$$

where

ω_η undamped natural frequency
ζ damping ratio

By examining equations (5.12) and (5.13), it may be seen that

$$2\zeta\omega_\eta=\frac{K_3}{J_o} \quad \text{and} \quad \omega_\eta^2=\frac{K_1K_2A}{J_o} \tag{5.14}$$

or by combining and simplifying

$$\omega_\eta=\sqrt{\frac{K_1K_2A}{J_o}} \quad \text{and} \quad \zeta=\frac{K_3}{2\sqrt{K_1K_2AJ_o}} \tag{5.15}$$

This second-order differential equation is similar to that of a simple open-loop spring-mass-damper system as discussed in chapter 3 on "Gyroscopes." This equation is repeated here for comparison.

$$\frac{d^2\theta}{dt^2}+\frac{C}{J}\frac{d\theta}{dt}+\frac{K\theta}{J}=\omega_i \tag{5.16}$$

where

C viscous damping, output axis
J moment of inertia, output axis
K spring constant, output axis
ω_i turning rate about the input axis

From equations (5.13) and (5.16), the following familiar results are obtained:

$$\omega_\eta=\sqrt{\frac{K}{J}} \quad \text{and} \quad \zeta=\frac{C}{2\sqrt{KJ}} \tag{5.17}$$

Thus, it may be seen that a simple position servomechanism very closely resembles a damped spring-mass open-loop system in which the spring rate is proportional to the loop gain and the damping constant is proportional to the motor velocity constant.

A set of normalized curves describing the frequency response and transient response of equation (5.13) is shown in figures 5.4, 5.5, and 5.6. Although these curves represent an elementary second-order linear system, most higher order servo loops have two predominant "least-damped roots" and are sufficiently linear for these curves to approximate system behavior. These curves show that except for $\zeta=0$, the steady-state value eventually damps out to a value of 1. The curves for low values of ζ damp out more slowly than for higher ζ, with the optimum about 0.7 of critical damping ($\zeta=1$). It takes much longer to reach the final steady-state value when ζ becomes large. The damping ratio depends on several variables in the loop, largely on the motor velocity characteristics and the loop gain. An increase in damping can usually be achieved by lowering the loop gain. The speed of response of the system is a function of both ω_n and ζ. Increasing loop gain will increase ω_n but decrease ζ, so that if significant changes in response are required, the addition of networks or change of servo components is required.

STABILITY

There are various ways of determining the stability of a servomechanism, many of which are too complex for discussion in this elementary discussion. Most of the methods deal with the open-loop

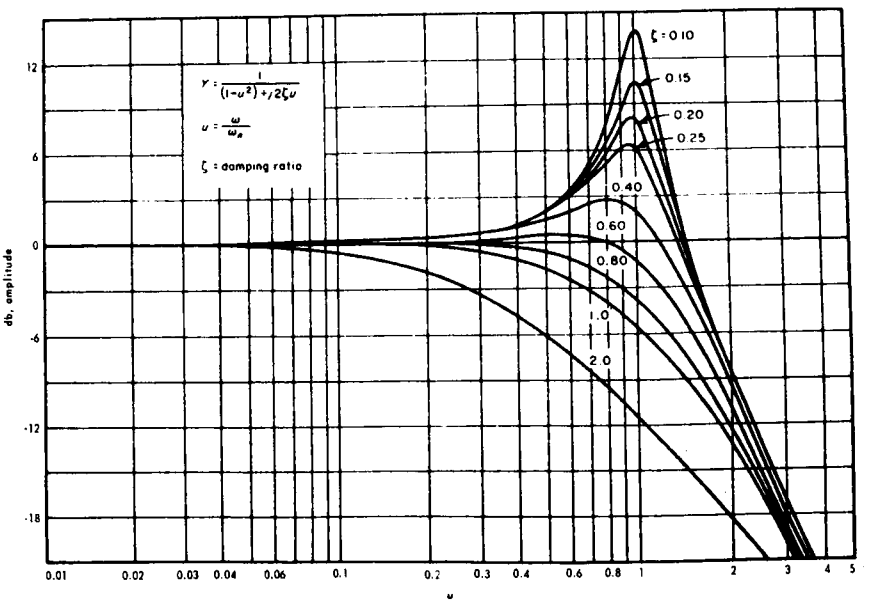


FIGURE 5.4—Amplitude of the second-order transfer function.

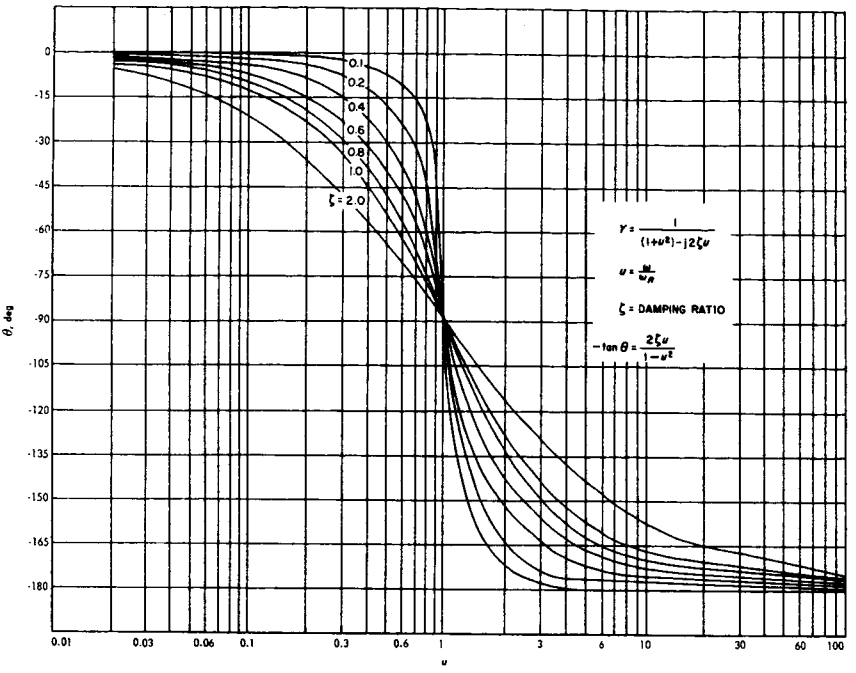


FIGURE 5.5—Phase shift for the second-order transfer function.

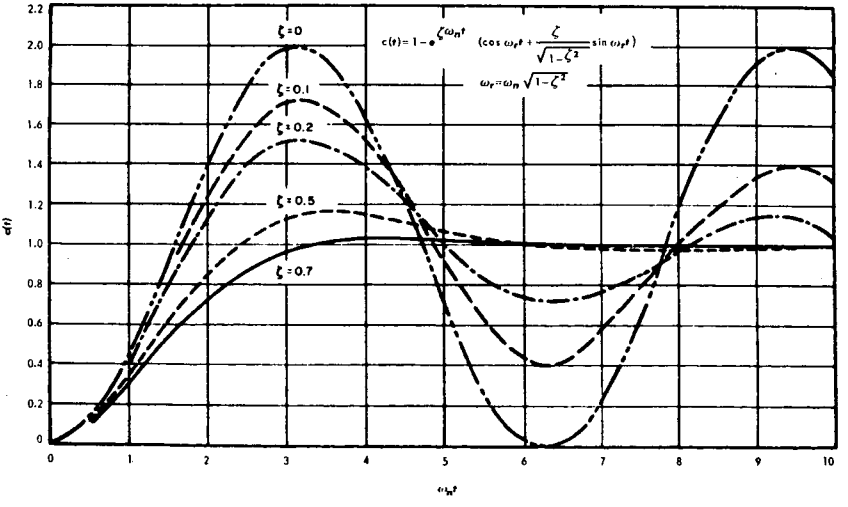


FIGURE 5.6—Transient response curves for various values of ζ.

transfer function of the servomechanism and some method for determining the optimum gain to incorporate in the closed loop. This can be accomplished by plots of the amplitude and phase response in linear or polar plots or by the behavior of the roots of the differen-

tial equation as a function of loop gain. The fundamental principle of all the techniques is to determine whether the gain of the loop is equal to or greater than unity at the time that the phase shift of the loop goes to 180° ($A\beta \rightarrow -1$). The root-locus technique plots the roots of the differential equation as a function of gain and determines when the sign of the roots goes positive. If the roots are positive, the transient response has the following form:

$$f(t) = e^{+\alpha t}(A \sin \omega t + B \cos \omega t) \tag{5.18}$$

Thus, the system would have unsatisfactory response, since equation (5.18) has the increasing oscillatory characteristics of an unstable system.

A plot of the open-loop phase and decibel gain versus the logarithm of frequency of a system is known as a Bode diagram. In the example shown in figure 5.3, if the motor time constant is 0.1 second, the Bode diagram of the loop is as shown in figure 5.7. The plot consists of a straight line of slope −6 dB/octave until a frequency of $\omega = 10$ is reached at which point the asymptote changes to −12 dB/octave. The curve is plotted for loop gain = 1. The phase curve starts out at 90° lag and becomes 180° lag corresponding to an arctangent curve. The degree of stability of a system using the

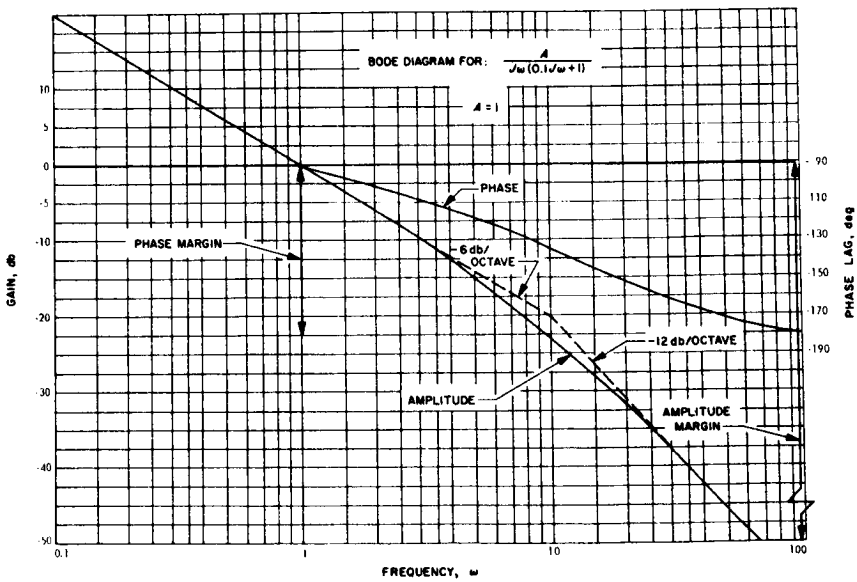


FIGURE 5.7—Bode diagram of a loop.

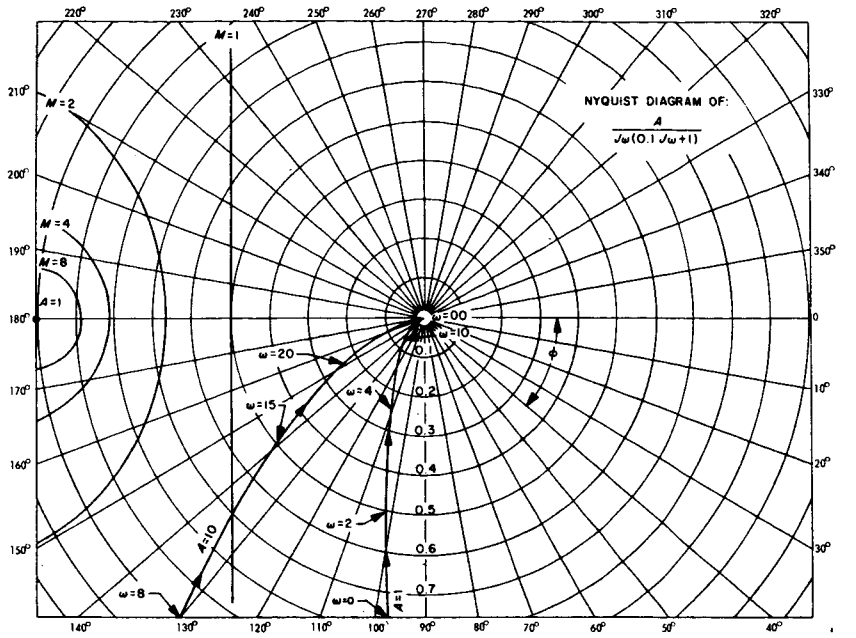


FIGURE 5.8—Nyquist diagram of a loop.

Bode diagram is determined by its gain or phase margin. From figure 5.7 it may be seen that at this gain the simple position servo of figure 5.3 is extremely stable, having a phase margin of 90° and an amplitude margin of greater than 50 dB. Actually, in this simple example of a second-order system, infinite gain is required to make the loop unstable.

A polar plot of the amplitude versus phase is known as a Nyquist diagram. The frequency is varied from zero to infinity at some specified value of gain and the behavior of the plot near the $\phi = -180°$, $A=1$ point is a criterion for determining stability. A Nyquist diagram of the system shown in figure 5.3 is shown in figure 5.8. On this Nyquist diagram may also be plotted circles of constant amplitude ratio (M circles) so that the resonant rise may be determined directly from the diagram.

The damping ratio, ζ, of a closed-loop second-order system may be determined from the relationships plotted in figures 5.9 and 5.10. A close approximation of ζ as a function of phase margin in the region $0 < \phi_m < 40°$ is given by:

$$\zeta = \frac{\pi}{360} \phi_m \tag{5.19}$$

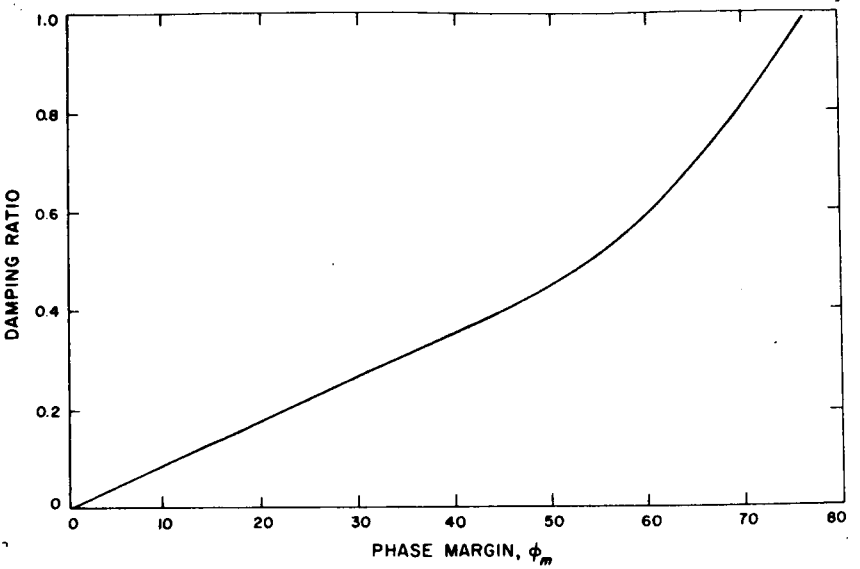


FIGURE 5.9—Damping ratio ζ versus phase margin ϕ_m.

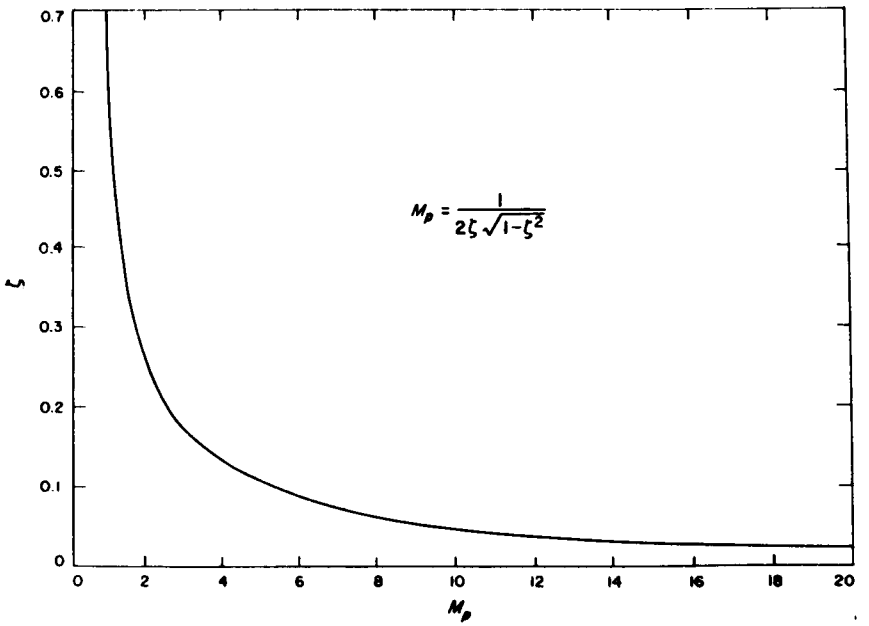


FIGURE 5.10—Damping ratio ζ versus maximum M, M_p.

The damping ratio of a second-order system is related to the amplitude ratio M by:

$$M=\frac{1}{2\zeta\sqrt{1-\zeta^2}} \tag{5.20}$$

Thus, for any value of gain, the damping ratio and hence the stability of a closed loop may be determined graphically. Although most servomechanisms are more complex than the second-order system discussed here, a close approximation of their characteristics may be made by comparison with second-order loops.

CHAPTER 6

Analog Computers

MOST TYPES OF GUIDANCE SYSTEMS use some form of computer to solve the guidance equations used to obtain steering and shutoff commands from the measurements of rocket or spacecraft motion. The computer may either be carried in the vehicle or located on the ground and connected to the vehicle by radio channels.

There are two main types of computers:

(1) *Analog computers* (ref. 9), in which physically measurable quantities in the computer, such as voltage or mechanical motion, are made to behave in ways analogous to the physical quantities described by the equations being solved. For example, a voltage in the computer might be made to behave as the analog of a velocity or position component of the spacecraft motion.

(2) *Digital computers*, which solve the required equations by numerical counting procedures and obtain numbers describing the quantities of the equations being solved. For example, the number representing a velocity component of the spacecraft motion may be increased periodically as a result of pulse inputs from a pulse-rebalanced force-balance accelerometer. (See ch. 4.) In a more complex example, the differential equation describing the motion of the spacecraft may be numerically integrated by some technique such as the Runge-Kutta method.

In this section, we shall first describe analog computers, and follow this by a description of digital computers in chapter 7.

ANALOG COMPUTER ELEMENTS

The most common analog computer elements are electronic. Some mechanical components are also used. The slide rule is a simple and familiar example of an analog computer.

If we wish to multiply a quantity A by another quantity B to solve the equation $C=AB$, we may either perform the multiplication directly (as we shall discuss later under Digital Computers), or we may recall that $\log C=\log A+\log B$ and, after determining $\log A$ and $\log B$ from tables, add them to obtain $\log C$. Then, by a second reference to the tables, we may determine C.

The slide rule eliminates much of the labor involved in this numerical process, as is shown in figure 6.1. Two pieces of wood or metal are graduated in such a way that the distance along each piece from a reference or index mark to a numerical graduation represents (or is the analog of) the logarithm of that number. Then, if the pieces are alined so that the index mark on the second piece is opposite the graduation corresponding to A on the first piece, the total distance from the index on the first piece to the graduation corresponding to B on the second piece is the sum of the length from the index to the graduation A on the first piece (corresponding to log A) and from the index to the graduation B on the second piece (corresponding to log B). Hence, the total length, as measured on the first piece, corresponds to log A+log B, and the number on the first piece opposite to the graduation B on the second piece is the numerical value of C.

We shall now turn our attention to some elements of electronic or electromechanical analog computers. The heart of the modern analog computer is the operational amplifier. This is a high-gain (several hundred thousand), high-input-impedance dc amplifier used as the forward element in an electronic servo loop. The operation of this loop is described below. Let us consider for the moment a single-input device. The amplifier is represented by the symbol in figure 6.2,

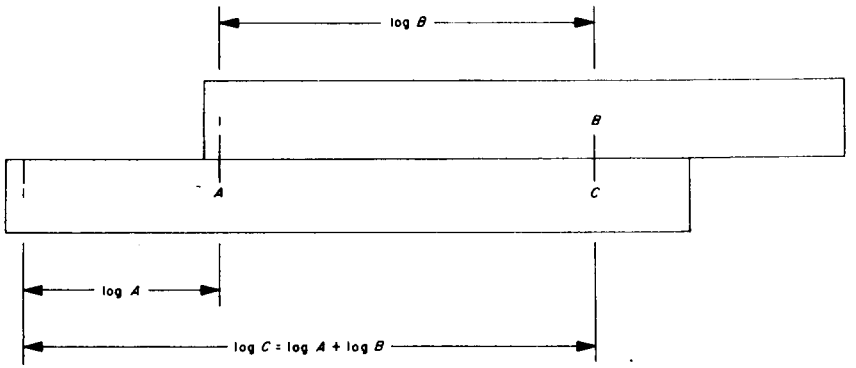


FIGURE 6.1—Operation of a slide rule.

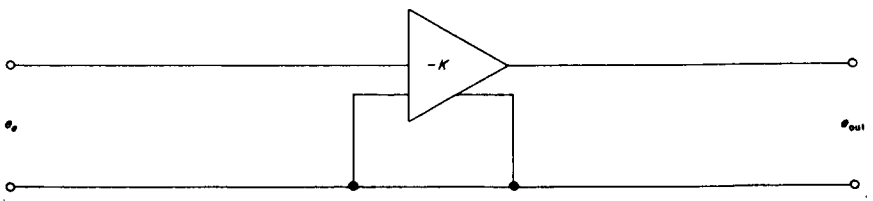


FIGURE 6.2—Operational amplifier.

where K denotes the magnitude of the gain from input to output. The outputs of all amplifiers are usually referenced to a common ground.

An approximate idea of the way in which the operational amplifier is used in analog computing may be derived by making the approximations (usually true in practice) that

(1) The input impedance of the amplifier is very high.
(2) The gain, K, is very large.
(3) The output impedance is very low.

In computing use, a voltage signal e_{in} is connected to the input terminals of the amplifier through an impedance Z_{in}, and the output of the amplifier is returned to the input through an impedance Z_{fb}, as shown in figure 6.3.

The properties of the amplifier are such that

$$e_{out} = -Ke_e$$

Hence

$$e_e = -\frac{e_{out}}{K}$$

Because the input impedance of the amplifier is assumed here to be infinite, all current flowing into the input node through Z_{in} must flow out through Z_{fb}, and hence, due to the sign convention

$$i_{in} = -i_{fb}$$

The input current is

$$i_{in} = \frac{e_{in} - e_e}{Z_{in}} = \frac{e_{in} + e_{out}/K}{Z_{in}}$$

and the output current i_{fb} is

$$i_{fb} = \frac{e_{out} - e_e}{Z_{fb}} = \frac{e_{out} + e_{out}/K}{Z_{fb}}$$

Hence

$$\frac{e_{in}}{Z_{in}} + \frac{e_{out}}{KZ_{in}} = -\frac{e_{out}}{Z_{fb}} - \frac{e_{out}}{KZ_{fb}}$$

and, rearranging

$$\frac{e_{out}}{Z_{fb}} = -\left[\frac{e_{in}}{Z_{in}} + \frac{e_{out}}{KZ_{in}} + \frac{Z_{in}}{Z_{fb}}\frac{e_{out}}{KZ_{in}}\right]$$

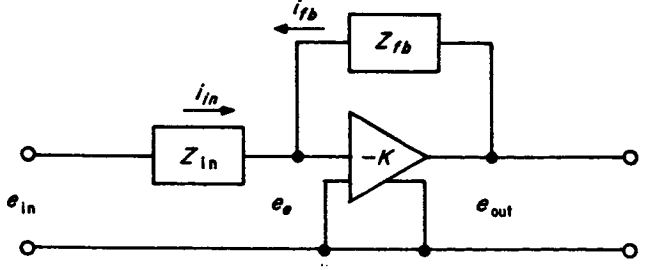


FIGURE 6.3—Operational amplifier with feedback.

or, collecting terms

$$e_{out}=\frac{-Z_{fb}}{Z_{in}}\left[e_{in}+\frac{1}{K}\left(1+\frac{Z_{in}}{Z_{fb}}\right)e_{out}\right]$$

In practice, e_{in} and e_{out} are similar in magnitude, as are Z_{in} and Z_{fb}. Hence, because K is large, the second term inside the brackets on the right side of the equation is very small and may be neglected. Thus

$$e_{out}\approx-\frac{Z_{fb}}{Z_{in}}e_{in}$$

Suppose that $Z_{fb}=R_2$ and $Z_{in}=R_1$. Then

$$e_{out}=\left(-\frac{R_2}{R_1}\right)e_{in}$$

and the configuration shown in figure 6.4 may be used to change the sign and level of voltage e_{in}.

If the gain, K, were to increase without limit, the voltage required at the input of the amplifier required to maintain an output would be vanishingly small, while the infinite input impedance would still require that $i_{fb}=-i_{in}$.

Now, if Z_{fb} is a capacitor, C, under the approximations stated above

$$e_{out}(t)-e_{out}(t_0)=\frac{1}{C}\int_{t_0}^{t}i_{fb}dt \tag{6.1}$$

But, if Z_{in} is a resistor of value R

$$i_{fb}=-i_{in}=\frac{-e_{in}}{R} \tag{6.2}$$

and thus

$$e_{out}(t)=e_{out}(t_0)-\frac{1}{RC}\int_{t_0}^{t}e_{in}dt \tag{6.3}$$

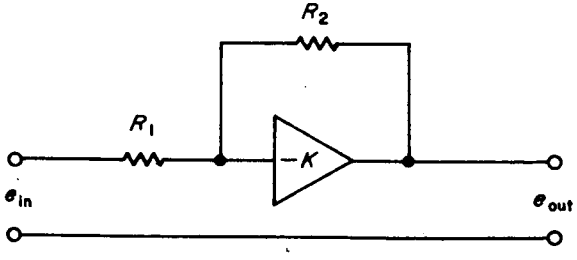


FIGURE 6.4—Configuration used to change sign and level of voltage e_{in}.

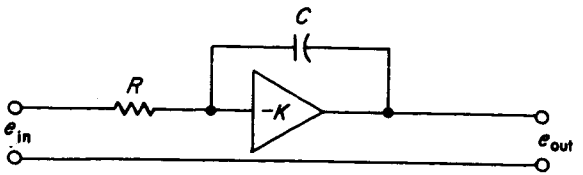


FIGURE 6.5—Configuration to integrate the signal e_{in}.

Thus, the configuration in figure 6.5 may be used to integrate the signal e_{in}, with the gain or the time scale being determined by the product RC.

Other combinations of elements may be inserted in the feedback path to provide other operations. The elements used in such applications are usually linear; that is, the parameters of the elements do not vary with voltage.

Let us now consider an amplifier with multiple inputs arranged as shown in figure 6.6.

Here

$$e_{out}=i_{fb}Z_{fb} \tag{6.4}$$

and

$$i_{fb}=-(i_1+i_2)=-\left(\frac{e_1}{Z_1}+\frac{e_2}{Z_2}\right) \tag{6.5}$$

Hence

$$e_{out}=\left[\frac{Z_{fb}}{Z_1}e_1+\frac{Z_{fb}}{Z_2}e_2\right] \tag{6.6}$$

The summation of inputs may, of course, be extended.

We now have developed the operation of enough components to show how an analog computer might be used to model the behavior of a single spring-mass-dashpot system.

The basic equation of motion states that

$$m\ddot{x}+c\dot{x}+Kx=f(t) \tag{6.7}$$

To simulate this equation, we first solve for the highest derivative

$$\ddot{x}=-\frac{c}{m}\dot{x}-\frac{K}{m}x+\frac{f(t)}{m} \tag{6.8}$$

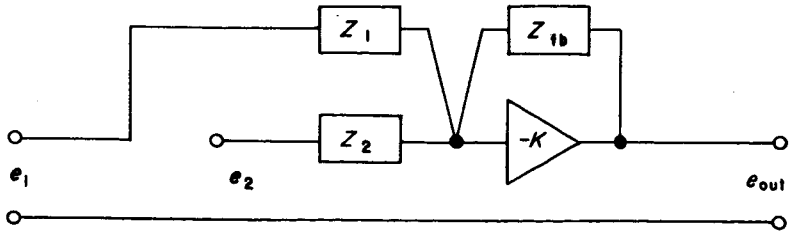


FIGURE 6.6—Amplifier with multiple inputs.

Now, if we let a voltage e_1 be the analog of $\ddot{x}$, and let this be the input to an integrator ($Z_{in}=R$, $Z_{fb}=C$), then the output e_2 of the integrator will be analogous to $\dot{x}$. If we then integrate this again, we shall obtain the analog of x.

We may then use these analog voltages as the inputs to a summing amplifier (several resistive inputs) to obtain as the output a voltage analogous to $\ddot{x}$. These manipulations are shown in figure 6.7.

The lower amplifier is necessary to obtain a voltage of the proper sign for $\dot{x}$, because each operational amplifier causes a sign inversion.

If the numbers involved are reasonable, the parameter values might be chosen to yield $R_1C_1=1$ second$=R_2C_2$, $R_3/R_4=1.0$, $R_0/R_5=c/m$, $R_0/R_6=K/m$ and $R_7/R_0=1.0$. One volt change in e_1 might represent 1 ft/sec^2 change in $\ddot{x}$; 1 volt in e_2, -1 ft/sec in $\dot{x}$; 1 volt in e_3, 1 ft in x; and 1 volt in e_4, $+1$ ft/sec in $\dot{x}$. The behavior of the system as a function of time might then be determined by recording the changes in e_3 and e_4 on a paper strip chart moving at constant speed under pens with deflections proportional to the values of e_3 and e_4 (i.e., 1 cm deflection per volt, which thus translates into 1 cm/ft and 1 cm/ft/sec for the e_3 and e_4 traces, respectively).

If the numbers are not convenient, different scale factors may be assigned to the voltages in question, and the parameter values adjusted until reasonable values are obtained. Discussion of this subject here is precluded by space limitations, but it may be found in standard texts on analog computation.

In the previous paragraphs, no explicit mention was made of the source of the driving function $f(t)$. If it were a single-step input, a switch to a fixed-voltage supply might be closed at the proper instant. Other methods are necessary for generating the terms of more complex functions. These methods are described below, together with other analog computer components not previously discussed.

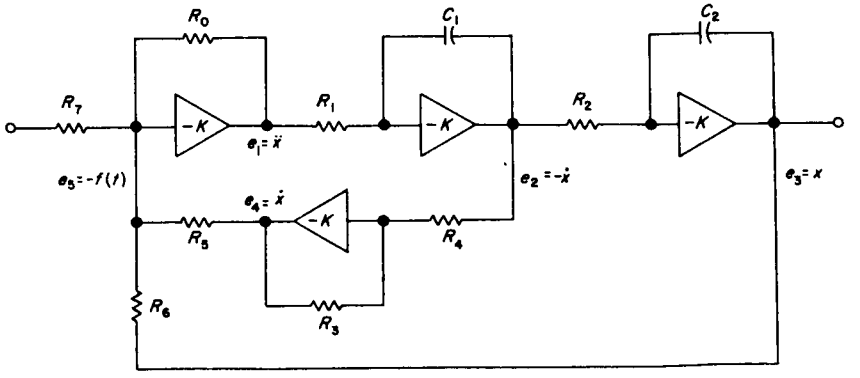


FIGURE 6.7—Analog of second-order system.

The potentiometer is used in many of these devices. It is simply a variable resistor, the variation being accomplished by means of a shaft rotation, as shown in figure 6.8.

The principle of use is as a voltage divider. If R denotes the total resistance, and $r(\theta)$ represents the resistance between one end and the movable arm, then if a voltage e_1 is impressed across the total resistance, the voltage e_2 between the end in question and the movable arm is

$$e_2 = e_1\left[\frac{r(\theta)}{R}\right] \tag{6.9}$$

The potentiometer principle is used in the construction of servo function generators, integrators, and multipliers. In its most common form, the resistance of the potentiometer increases in direct proportion to the displacement θ from one end, $r(\theta) = K\theta$. If such a potentiometer is excited across its outer terminals by a fixed voltage, then the voltage output from the movable arm is directly proportional to the shaft position.

A servomultiplier may be constructed as shown in figure 6.9.

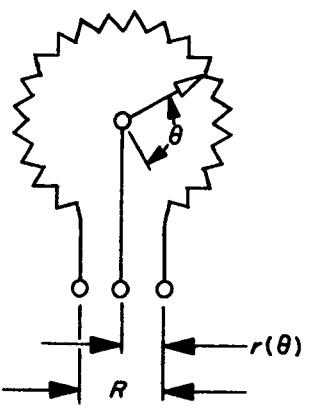


FIGURE 6.8—Potentiometer.

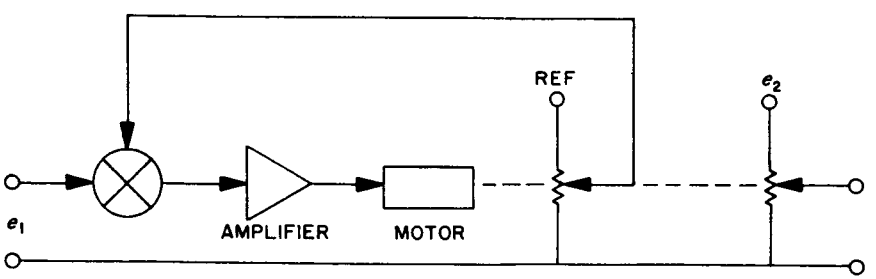


FIGURE 6.9—Servomultiplier.

The output of the first potentiometer is proportional to the shaft rotation, and the action of the servo loop thus forces the shaft rotation to be proportional to e_1. Then the output of the second potentiometer is proportional to its excitation voltage e_2 and the shaft rotation, which in turn is proportional to e_1. Hence, the output of the second potentiometer is proportional to the product of e_1 and e_2.

A servo integrator is made by using the configuration shown in figure 6.10.

A tachometer on the motor shaft generates a voltage proportional to shaft rate. The servo action forces this voltage to be equal to e_1, and thus forces the shaft position and potentiometer output to be proportional to the time integral of e_1.

Potentiometers need not be made in such a way that resistance variations are uniform; for example, the variation in resistance may be made proportional to the sine or cosine of shaft rotation, and multiplication by trigonometric functions as required in vector coordinate transformations may be obtained.

By proper connections, a servo may also be made to divide one function by another.

In guidance applications, the basic function of the analog computer may be to integrate measured accelerations and to command shutoff of the rocket motor when the integral reaches a certain value. A relay amplifier is used here to sense when the integrator output exceeds a specified value and changes the state of a relay when this condition is fulfilled. Relay amplifiers also find many uses in simulation. Another function often required in simulation is that of limiting the excursions of a particular voltage to simulate the effects of limits of travel, and so on.

The accuracy of analog computation is limited by the accuracy and stability of the input and feedback impedances, stability of the reference voltage sources, and drift and finite gain of the operational amplifiers. Typical accuracy of a single high-quality operational amplifier in an analog simulator is on the order of 0.1 percent of full scale. High-quality analog integrators for guidance purposes reach the limits of practicality at about 0.03 percent.

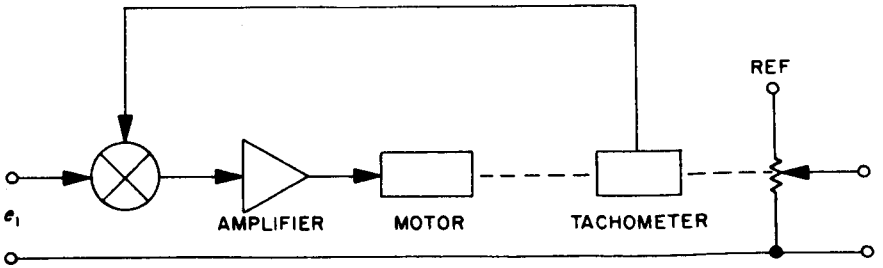


FIGURE 6.10—Servo integrator.

CHAPTER 7

Digital Computers

THE LIMITS OF ACCURACY stated for analog computers are not adequate for many of the design and guidance computations associated with a modern spacecraft. To execute such computations, a different approach—digital computation—is used. Also, certain types of equations are very difficult to solve with analog computers.

In the analog computer, quantities that are easy to manipulate, such as voltages and shaft rotations, are made to behave in a way that is the analog of the behavior of the actual physical quantity, such as velocity, being studied. The behavior of the quantity is determined by measuring the behavior of its analog at the particular time of interest.

In digital computation, mathematical expressions are constructed to describe the behavior of the physical quantities being studied. These equations are then solved numerically, and the numerical values of the variables are investigated to determine the behavior of the physical quantities. In theory, the accuracy of this process is basically limited by the number of "significant figures" carried in the computations.

ADDING MACHINES AND CALCULATORS

Suppose that we wish to compute $c=a+b$. The example of the digital computer most familiar to us is probably the adding machine, or its big brother, the desk calculator. The number representing a is entered into the machine by positioning wheels according to the digits in a. Next, the positions of these wheels are incremented by the digits in b, with proper account being taken of carries. The resulting wheel positions represent the digits of c.

The speed of such machines is limited by the mechanical elements. In some problems, such as the integration of nonlinear differential equations, considerably more speed is desirable. It can be obtained by using electronic devices. Further limitations on the speed of problem solving with the calculator are the need for a human operator to insert numbers into the machine and push the proper buttons to

accomplish the basic operations of addition and shifting, combined with the simple programed combinations of these operations: subtraction (addition with reversed sign); multiplication (repeated addition, shortened by shifting); division (repeated subtraction, shortened by shifting); and on some machines transferring a number from one set of dials to another (for continued products, $d=abc$), squaring (a particular form of multiplication), and square-root extraction (a special form of repeated subtraction and shifting).

This speed limitation can be removed by providing electronic storage for each of the numbers and instructions to be used, for transfer of the information from one location to another, and control of the operations to be performed.

PROBLEM-SOLVING PROCESS

Before discussing how these tasks are accomplished, we shall discuss the nature of the tasks in some detail, using as an example the solution of a quadratic equation

$$ax^2+bx+c=0 \tag{7.1}$$

The computer is to be used to determine values for x that satisfy the equation for given values of the parameters a, b, and c. The steps in accomplishing this task are listed below:

(1) Receive from the user the *numerical* values of a, b, and c for which corresponding *numerical* values of x are desired and such data as are necessary to identify the set of parameters received (see item (3)).

(2) Solve the equation using the quadratic formula

$$x=\frac{-b\pm\sqrt{b^2-4ac}}{2a} \tag{7.2}$$

For the present, we shall assume that the computer is capable of the same operations as a desk calculator, i.e., adding, subtracting, multiplying, and dividing, as well as square-root extraction, and we shall write out in detail the steps necessary to accomplish the task with such a machine. This list is shown in table 7.1.

(3) Supply to the user the *numerical* values of x which satisfy the equation for the given numerical values of the parameters a, b, and c, together with sufficient information to identify the sets of parameters used.

As table 7.1 shows, there are three major parts to the process of solving the problem:

(1) INPUT of the *specific values* of the necessary parameters and identifying data.

(2) SOLUTION of the problem by performing the necessary mathematical operations upon the input data.

(3) OUTPUT of the *specific values* of the variables that are valid solutions of the equation for the specific values of the parameters input, together with identifying data defining the set of parameters used.

It is worth taking special note of the order and dependence of these steps on one another. Without INPUT the computer performs no useful function because no demands have been made upon it. Once INPUT is received, the computer solves the problem defined by the specific values it receives. If the input is in error, so are the results, although the computer solved the problem "correctly" for the erroneous input data. Finally, OUTPUT must take place before the computer's solution to the problem yields value by being placed in the hands of the user.

The input/output process is the interface between the computer and the user and is the locus of many problems in the use of a digital computer.

The detailed program of operations is listed in table 7.1. For concreteness, it is assumed that the communication between the user of the results and the computer operator is verbal and that the operator has a scratch pad with numbered lines.

The human operator, in executing this procedure, might keep the list or program under his left hand and keep his index finger on the line corresponding to the step being executed. After the execution of a step, the index finger would be advanced to the next line, which has a number one higher than the previous line, and the new step indicated would be executed.

As stated previously, the limitations on speed of the desk calculator are as follows:

(1) The necessity of controlling the machine with a human operator who writes down input data, reads the program of instructions, punches keys to enter the proper numbers, punches buttons to control the operation of the machine, writes down intermediate and final results (in some machines this function is performed by the machine which prints the results of its calculations on adding-machine paper), and reads the final results to the user.

(2) The speed with which the mechanical linkages and wheels operate to perform the necessary counting operations.

FUNCTIONAL NATURE OF THE TASK

It was stated previously that the speed of operation can be improved if some of these tasks are performed electronically. To see how this

238-635 O—67——5

TABLE 7.1—Detailed Program of Operations

1. Ask the user for a new set of input data.
2. Write on line 1 the first number (identifying the set of values a, b, and c).
3. Write on line 2 the second number (a).
4. Write on line 3 the third number (b).
5. Write on line 4 the fourth number (c).
6. Enter as a multiplicand the number from line 3(b).
7. Enter as a multiplier the number from line 3(b).
8. Clear and multiply.
9. Write on line 5 the result (b^2).
10. Enter as a multiplicand the number from line 2(a).
11. Enter as a multiplier the number 4.0.
12. Clear and multiply ($4a$).
13. Enter as a multiplicand the result.
14. Enter as a multiplier the number from line 4(c).
15. Clear and multiply.
16. Write on line 6 the result ($4ac$).
17. Clear the machine.
18. Add the number from line 5(b^2).
19. Subtract the number from line 6($4ac$).
20. Write on line 7 the result (b^2-4ac).
21. Check the sign of the result: if (−), go to step 22; if (+) or zero, go to step 23.
22. Reverse the sign of the result in the calculator.
23. Extract the square root of the result in the calculator ($|b^2-4ac|$).
24. Write the result on line 8 ($\sqrt{|b^2-4ac|}$).
25. Enter as a multiplicand the number from line 2(a).
26. Enter as a multiplier the number 2.0.
27. Clear and multiply.
28. Write on line 9 the result ($2a$).
29. Enter as a dividend the number from line 8 ($\sqrt{|b^2-4ac|}$).
30. Enter as a divisor the number from line 9 ($2a$).
31. Divide.
32. Write on line 10 the result ($\sqrt{|b^2-4ac|}/2a$).
33. Enter as a dividend the number from line 3(b).
34. Enter as a divisor the number from line 9 ($2a$).
35. Divide.
36. Reverse the sign of the result.
37. Write on line 11 the result ($-b/2a$).
38. Check the sign of the number on line 7. If (−), go to step 49; if (+) or zero, go to step 39.
39. Clear the machine.
40. Add the number from line 11 ($-b/2a$).
41. Add the number from line 10 ($\sqrt{|b^2-4ac|}/2a$).
42. Write on line 12 the result ($[-b+\sqrt{|b^2-4ac|}]/2a$).
43. Write on line 13 the number 0.0.
44. Subtract the number from line 10.
45. Subtract the number from line 10.
46. Write on line 14 the result ($[-b-\sqrt{|b^2-4ac|}]$).
47. Write on line 15 the number 0.0.
48. Go to step 54.
49. Write on line 12 the number from line 11 ($-b/2a$).
50. Write on line 13 the number from line 10 ($\sqrt{|b^2-4ac|}/2a$).
51. Write on line 14 the number from line 11 ($-b/2a$).
52. Reverse the sign of the number from line 10.
53. Write on line 15 the result ($-\sqrt{|b^2-4ac|}/2a$).
54. Tell the user results are available.
55. Read the number on line 1 (identifying number).
56. Read the number on line 12 (real part of 1st root).
57. Read the number on line 13 (imaginary part of 1st root).
58. Read the number on line 14 (real part of 2d root).
59. Read the number on line 15 (imaginary part of 2d root).
60. Go to step 1 (ready for next problem).

is done, let us first inspect the functional nature of the task. The functions include:

(1) Receiving input
(2) Storing information
    (*a*) Program of instructions
    (*b*) Input
    (*c*) Intermediate and final results
(3) Executing arithmetic operations
(4) Transmitting output
(5) Controlling the sequence of operations

These functions will now be described in more detail.

The input and output functions consist of the following chain:

(1) Receiving data in one physical form and format from a particular location
(2) Changing the form and format of the data
(3) Transmitting the changed form and format to a new location

For instance, in the input function described in "Problem-Solving Process" (p. 58), the data are *received* from the *user* in *verbal* form in a certain *order*, *changed* to *written symbols* by the operator, and *written* on the *computation pad* on successive *numbered lines*. The reverse process is followed for output.

The information is stored as written symbols on paper in ordered lists. One list serves as the program of instructions and another list serves as the input data, intermediate results, and final results. The first list is prepared before the problem starts, while the second list is filled out during the course of the problem. At the present time, it is merely noted that many symbols are used to represent the numbers (a total of 13—the 10 digits, decimal point, +, and −), and many more are used to represent the program of instructions (in addition to the previous 13, the 26 letters of the alphabet and blank spaces, as well as the comma, semicolon, right parenthesis, and left parenthesis, making 44 in all). This subject is discussed in greater detail later.

The arithmetic operations are executed by the desk calculator, which is called the *arithmetic unit* for the remainder of the discussion. According to control instructions transmitted to it through the punching of buttons, the arithmetic unit will add, subtract, multiply, and divide. According to the way in which numbers are entered with keys and buttons, the numbers will serve as addend, augend, subtrahend, minuend, multiplier, multiplicand, divisor, and dividend. The results are displayed on rows of dials, which also serve as intermediate storage during the repeated-addition operation of multiplication, and the repeated-subtraction operation of division.

The control of the operations is provided by the human operator, who follows the ordered list of the program of instructions and punches

the proper keys and buttons to transmit information from the storage (papers containing the lists) to the arithmetic unit, and vice versa.

MEMORY AND STORAGE

Consider the way a word is stored on a piece of paper: as an ordered row of symbols. A common office typewriter is capable of placing any one of about 38 symbols (or, if capitals and lowercase are considered, 64 symbols) on a piece of paper in addition to leaving blanks. In the writing of engineering text, about 60 more are used, including the Greek alphabet and such special symbols as the integral sign, and the partial-differential symbol. In addition, the arrangement of symbols on the line or as subscripts and superscripts also has meaning; for example, abc is different from $a_b{}^c$.

The key to increasing the speed of the digital computation process with electronic devices is to store the instructions and the numerical data electronically so that the information can be transmitted or operated upon quickly, often in intervals measured in microseconds or less. Typical digital computers for scientific computations and data processing recognize from 50 to 64 different symbols for input/output. However, if it were necessary to store each of these symbols differently in the machine, for example, as different voltage levels, one per symbol, the delicacy of design and adjustment necessary for successful operation would be very great. To avoid electronic problems and to minimize the need for delicate adjustment, ways have been devised to represent information electronically with just two states for each piece of equipment, corresponding to the extremes "high or low," "positive or negative," or "on or off."

Devices which are capable of only two states are referred to as binary devices. The characteristic of being always in one of two states permits many arithmetic and logical operations to be performed with ease by combinations of binary devices that possess one additional characteristic: the outputs of the device are determinable functions of the inputs. In addition, many electronic binary devices possess a third characteristic: once the device has been placed in either of its stable binary states according to the inputs it has received, the device will remain in that state until new inputs are received. Such devices are said to have memory. Devices possessing such characteristics include ferromagnetic cores and the electronic circuits known as flip-flops. Magnetic tape also possesses these characteristics as do certain relay circuits, some types of phosphors used in certain types of cathode-ray tubes (within the decay time), and certain other electronic and mechanical devices. Many circuits using electron tubes, transistors, and diodes possess the first two characteristics.

In discussing the storage of information represented in binary form, let us begin by considering the way in which numerical information is represented. Suppose that we say a particular item is 473.25 meters long. This is the equivalent to saying that the length of the item is represented by the following sum of lengths in meters: $4\times10^2+7\times10^1+3\times10^0+2\times10^{-1}+5\times10^{-2}$. Inspection of the number 473.25 and the sum reveals that the number of places a digit lies to the left of the place immediately preceding the decimal point is equivalent to the exponent of the power of 10 attached to that digit. In a similar way, we could express the same unit of length in terms of powers of 2. However, note that 2×2^n is equivalent to the sum $(1\times2^{n+1}+0\times2^n)$, so that we shall require only the two digits 0 and 1 to represent the number powers of 2. The number discussed above is equivalent to the sum

$$(1\times2^8+1\times2^7+1\times2^6+0\times2^5+1\times2^4+1\times2^3 +0\times2^2+0\times2^1+1\times2^0+0\times2^{-1}+1\times2^{-2}) \quad (7.3)$$

as can be verified by the reader with simple addition. Following our previous rule for notation, we can use the shorthand symbol 111011001.01_2 to denote the number 473.25_{10}, where the subscript is used to denote the base to which the shorthand refers.

If we have a string of binary devices, we can set them to the appropriate state to represent the number, as shown in figure 7.1. To be able to read the correct number out of these devices, we must be careful to sample their outputs in the proper order and recall the location of the "binary point."

A string of binary devices used to represent a piece of information is called a register, and the piece of information in it is called, for some obscure reason, a "word," rather than a "number." The number of binary information digits (once abbreviated "binits," but now shortened to "bits") contained in the word is equal to the number of binary devices. These outputs may either be sampled one after the other in the proper order and the information transmitted on a single line called "serial readout," or all of the outputs may be sampled together on separate lines called "parallel readout." Sup-

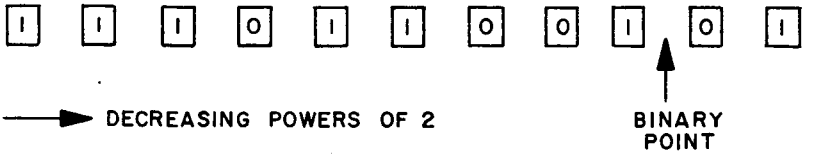


FIGURE 7.1—Register consisting of a string of binary devices.

pose that we let a positive voltage represent a 1 and a negative voltage represent a 0. Then we have the states of affairs shown in figure 7.2 for serial readout and figure 7.3 for parallel readout.

From the preceding discussion we can draw the conclusion that the storage device or memory of a digital computer is composed of registers with the memory characteristic. For ease of design, each memory word is of a fixed length. For example, in a large IBM computer there are 32 768 ($=2^{15}$) memory registers or cells, each with a capacity of 36 bits (binary information digits). Because there are exactly 2^{15} cells, we see that a 15-bit binary number can serve as the unique "address" of each memory register or cell.

So far, it has been shown how a decimal number can be represented as a binary number. However, the fact has also been introduced that most memory cells are of fixed length. If the location of the binary point is also fixed, we are restricted to dealing with numbers in a certain range if we are to be able to store them in these fixed-length cells. The processing of problems entirely in such *fixed-point* arithmetic is inconvenient, so a different method has been devised called *floating-point* arithmetic. For example, in decimal notation 473.25 is equal to 0.47325×10^3, hence a wide range of numbers may be represented with a small number of digits by limiting the number of significant digits and hence the accuracy one is willing to deal with. For instance, most desk calculators are limited to operations with 10 significant digits, which is sufficient for most operations. In the floating-point system, two numbers are used: the "normalized"

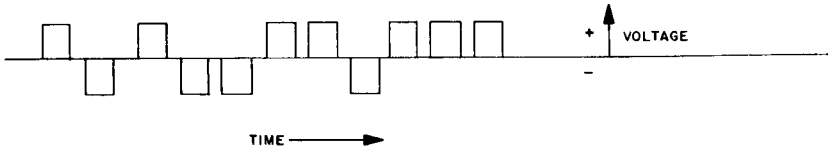


FIGURE 7.2—Voltage waveform representing serial readout, least significant bit first.

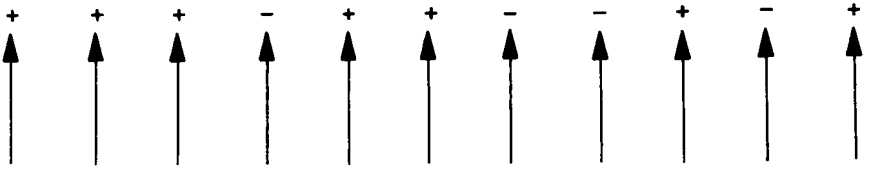


FIGURE 7.3—Voltage levels on parallel readout lines.

digits with the first digit immediately to the right of the point and the exponent to which the base must be raised to represent the number. For instance, we might represent 473.25 as 0.47325 $E+03$, meaning that 0.47325 is to be multiplied by 10^3; i.e., that the decimal point is to be moved 3 places to the right. Likewise, 0.00047325 can be represented as 0.47325 $E-03$.

A typical large IBM computer can handle 8 significant digits of positive and negative numbers lying between 10^{-38} and 10^{+38} in magnitude with its 36-bit memory cells. One bit is used for the sign, 8 bits are used for the binary exponent (converted to a number that is always positive by adding a positive constant), and the remaining 27 bits are used for the normalized digits. The storage of a floating-point number is an example of the use of a single memory word to store (i.e., represent) three more-or-less distinct pieces of information in the same way that the word "nonconformist" conveys three such pieces of information—the basic action, conforming; the indication that a person is involved, -ist; and the negation, non-.

The representation of instructions in memory cells will now be discussed. Inspection of the program of instructions for solving a quadratic equation indicated that there are 17 kinds of instructions involved, as listed in table 7.2.

Note that each instruction (table 7.2) refers either to the arithmetic unit loaded with numbers in the proper places (registers, by our previous definition) or to one or more other locations such as instruction-storage registers.

Although no instruction is less than 17 symbols long, including blank spaces, and although 42 different characters (all numbers, letters except J, 10 decimal digits, +, −, comma, semicolon, (,), and blank space) are required, each instruction could be uniquely identified by a group of not more than 5 numbers:

(1) A number from 1 to 17 indicating the instruction
(2) A number indicating the memory cell involved
(3) Up to three numbers indicating positions on the instruction program or indicating the source or sink of data

As the previous discussion has shown, several numbers can be placed in a single-memory word if care is taken to use the right digits in the word.

For example, the user may be denoted by the decimal digits 01, the location of an instruction in the program by the number attached to it in the list, and the storage location for a number by the line number stated in the program with 70 added to it (the output of the arithmetic unit is designated by the decimal digits 99). Missing numbers will be denoted by 00. Then the segment of the program beginning with instruction 9 would read as shown in table 7.3. Each

TABLE 7.2—Program of Instructions for Solving a Quadratic Equation

1. Ask _____ for information
2. Write at location _____ the number at _____
3. Enter as a multiplicand the number at location _____
4. Enter as a multiplier the number at location _____
5. Clear and multiply the multiplicand by the multiplier
6. Clear the machine
7. Add the number at location _____
8. Subtract the number at location _____
9. Check the sign of the number at location _____; if (−), go to instruction _____; if the number is zero, go to instruction _____; if (+), go to instruction _____
10. Reverse the sign of the number at location _____
11. Extract the square root of the absolute value of the number at location _____
12. Enter as a dividend the number at location _____
13. Enter as a divisor the number at location _____
14. Divide the dividend by the divisor
15. Go to instruction _____
16. Tell _____ that information is going to be read out
17. Read the number at location _____ to _____

TABLE 7.3—Segment of the Program

9.	02	75	99	00	00	(99 denotes the output of the arithmetic unit)
10.	03	72	00	00	00	
11.	04	86	00	00	00	(The number 4.0 must be written on line 16 before starting the program)
12.	05	00	00	00	00	
13.	03	99	00	00	00	
14.	04	74	00	00	00	
15.	05	00	00	00	00	
16.	02	76	99	00	00	
17.	06	00	00	00	00	
18.	07	75	00	00	00	
19.	08	76	00	00	00	
20.	02	77	99	00	00	
21.	09	99	22	23	23	
22.	10	99	00	00	00	
23.	11	99	00	00	00	
24.	02	78	99	00	00	etc.

of the two-digit groups could be changed to its binary form and the result stored in a register. For example, instruction 21 can be expressed in the 35-bit form:

21. 0001001 1100011 0010110 0010111 0010111

We thus see that a limited vocabulary of instructions can be stored efficiently if we give each instruction word a number and use that number for storage purposes. Other parts of the memory word can be used for location and instruction numbers associated with that instruction.

In some cases, it becomes necessary to store words not contained in the instruction vocabulary of the computer in the computer memory. The most common example is the storage of heading data to be printed out to identify the numbers on the printed computer output.

As was mentioned earlier, the common office typewriter is capable of about 39 (including the blank) symbols if we do not differentiate between capitals and lowercase letters. The same scheme can be followed as that used for the instructions, a number assigned to each of the characters to be used, and then the set of digits corresponding to the characters can be stored. A moment's reflection on the capabilities of binary numbers reveals that 6 bits are adequate to identify 64 different characters. Reserving the numbers 0–9 for the decimal digits and letting 10 correspond to a blank, the letters of the alphabet can then commence so that $11=A$, $12=B$, $13=C$, etc. Then the decimal numerical representation for IN BAD is 19 24 10 12 11 14, which would be stored in binary form as 010011 011000 001010 001100 001011 001110.

This example completes the discussion of the storage of information in binary form. It has been shown how decimal numbers can be converted to binary numbers, how floating-point numbers are used, and how a bounded set of symbols is ordered into a list and given corresponding numbers, which are then stored. It has also been shown how several numbers may be packed into a single word by proper placement and, thus, how both instructions and strings of symbols, each requiring several multibit words, can be stored efficiently.

INFORMATION PROCESSING

Processing of digital information is performed by devices with binary outputs that take on certain states according to the binary states of the inputs. Some of the simplest of these devices is the family of devices called gates, in which a single output depends on the states of two or more inputs. The first to be considered are two-input gates. Any gate can be described by a state table specifying the output state for each combination of input states. Consider state tables shown in figures 7.4, 7.5, and 7.6. The numbers to the left of the table indicate the state of the input coming from the left, the numbers at the top indicate the state of the input coming from the

top, and the numbers in the body of the table indicate the state of the output. Using the convention that

1="on," "high," "positive"
0="off," "low," "negative"

then the type of gate called the AND gate has the state table of figure 7.4.

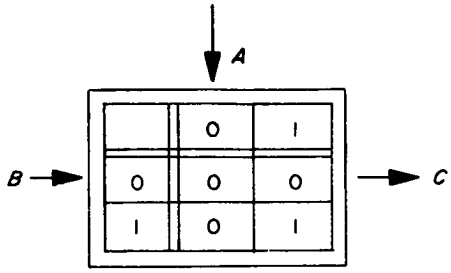


FIGURE 7.4—AND gate.

While the type of gate called the OR gate has the state table of figure 7.5.

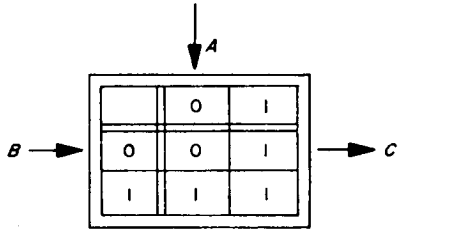


FIGURE 7.5—OR gate.

The so-called EXCLUSIVE OR gate has the state table of figure 7.6. This type of state table can be extended for gates with more than two inputs. For example, three-input OR gates and AND gates would have the state tables shown in tables 7.4 and 7.5, respectively.

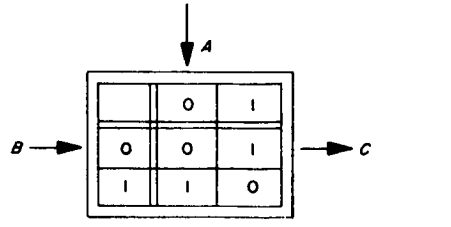


FIGURE 7.6—Exclusive OR gate.

TABLE 7.4—3-Input OR-Gate State Table

A	B	C	Output
0	0	0	0
0	0	1	1
0	1	0	1
0	1	1	1
1	0	0	1
1	0	1	1
1	1	0	1
1	1	1	1

OR gate

TABLE 7.5—3-Input AND-Gate State Table

A	B	C	Output
0	0	0	0
0	0	1	0
0	1	0	0
0	1	1	0
1	0	0	0
1	0	1	0
1	1	0	0
1	1	1	1

AND gate

Two single-input devices must also be mentioned: The inverter, with output state opposite to input state, and the unit delay, which puts out a pulse of the same state as the input one time unit after the input is received. Such devices are required in several operations.

Next, an example will be shown of the information flow in a device to produce a waveform that is the serial binary representation of the arithmetic sum of two numbers represented in serial binary form by input waveforms beginning at the same time. In each case, the least-significant bit is transmitted first. The state table for addition of a single digit is given in table 7.6.

The simplest gates are those which perform the AND and OR functions. This addition function will be shown for a single digit using these elements plus inverters.

The combination shown in figure 7.7 is called a half-adder. To handle carries, two half-adders are used together with an OR gate and a unit delay, as shown in figure 7.8.

TABLE 7.6—State Table for Addition of a Single Digit

A	B	$A+B$	Carry
0	0	0	0
0	1	1	0
1	0	1	0
1	1	0	1

Addition of single digit

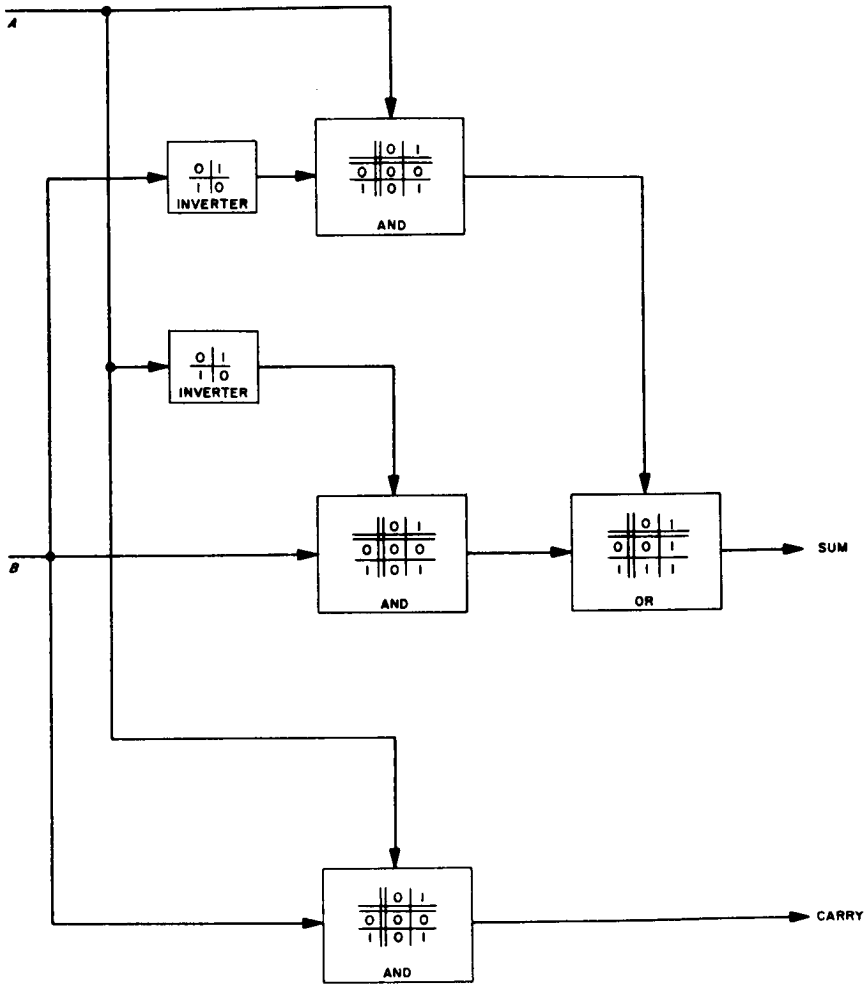


FIGURE 7.7—Half-adder.

The unit delay and the second half-adder causes the carry to be added to the sum of the next-highest pair of significant digits (because the least-significant bits enter first). If a carry results from this addition, it enters the OR gate and the unit delay causes this carry to be added to the next-highest pair, and so on. It appears from an initial inspection that the operation would be faulty if both a carry and a propagated carry occurred at the same time because the OR gate does not add them in the arithmetic sense. However, this state cannot occur as is shown below.

Consider the nth digit pair. Inspection of table 7.6 shows that if a carry is present, the digit sum must be 0. Thus, because one of

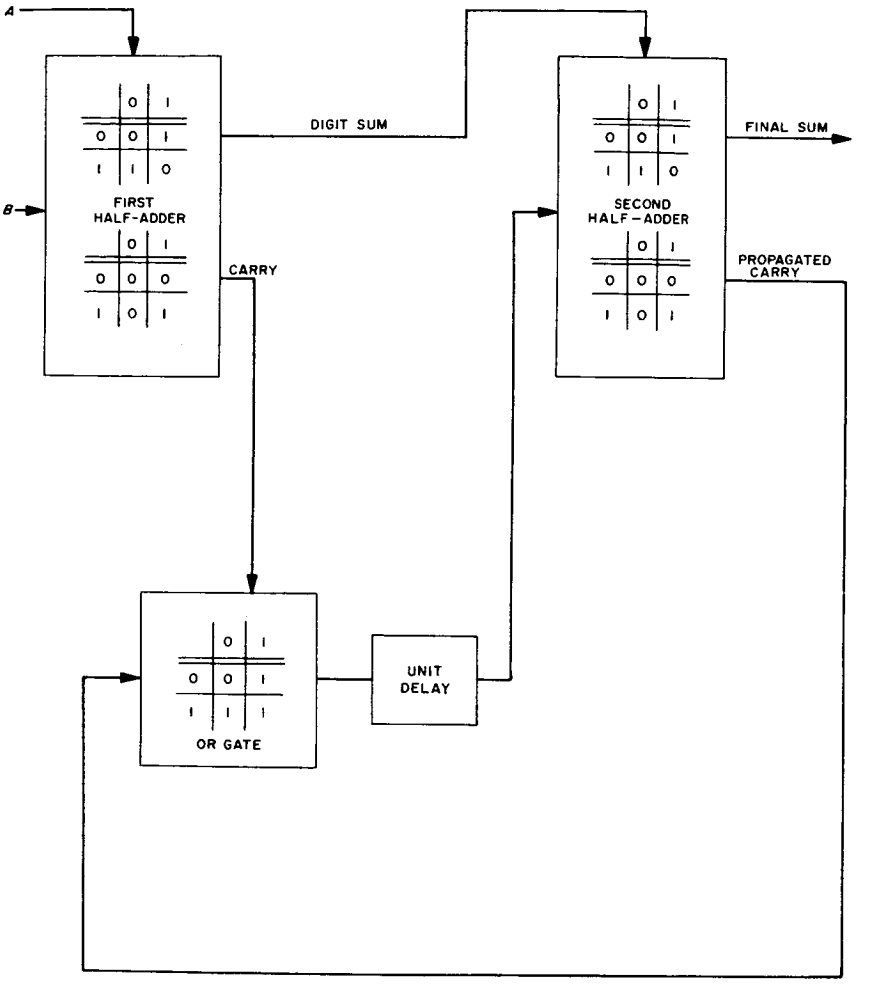


FIGURE 7.8—Full adder.

the inputs (digit sum) to the second half-adder is 0, a propagated carry cannot occur at the nth pair if a carry occurs, even though a delayed carry may be inserted into the second half-adder from the $n-1$st interval.

The preceding paragraphs have shown how binary devices may be connected to perform the operation of arithmetic addition on pulse trains that are serial representations of binary numbers. These devices may also be combined in other ways to perform the remaining operations required for digital computer operations. Standard texts on digital computers discuss these arrangements in detail.

CONTROL

The control unit contains a central clock to control the timing of the operations and to supply the necessary pulse trains for use by the various computing elements. The control unit also contains equipment to decode the instructions stored as binary numbers in the instruction storage registers and cause the arithmetic unit and the gates controlling transfers to and from the memory to perform the operations called for by the stored instructions. The operations are stored in numerically ascending sequence in the memory, and a register called the instruction counter in the control unit increases its count by one unit each time an instruction is retrieved from the storage to the control unit and executed. Thus, if the proper initial condition is inserted, the number in the instruction counter corresponds to the address of the next instruction to be retrieved. Instructions of the transfer type, such as "Go to instruction _____," or "Go to instruction _____ if _____ (a certain condition is met)," provide for branch or decision points in the program by modifying the contents of the instruction counter. The absolute "Go to" type of instruction is used to return a program to the beginning after a computation has been completed or to jump over a block of instructions in the program. The conditional type of transfer, dependent on whether or not a certain condition is met, may be used to call for different processes depending on the sign or magnitude of one of the intermediate results, or to call for or omit certain optional computing processes dependent on the value of an input quantity, thus providing operator control. These input quantities may either be read in with the rest of the data, or may be switch states controllable from the console that are tested each time the instruction is encountered.

A few words should be said about subroutines. Suppose that a certain process that takes many instructions, such as extracting the square root of a number, is used several times in a program. It is not economical of programing time or of computer memory space to write all of the instructions each time. Instead, each time a number is to have its square root extracted, that number may be placed in a particular location and the computer told to save the address of the next instruction in a special register. The computer is then instructed to transfer to the instruction beginning the square-root process, which instructs the computer to take the square root of the number in the aforementioned particular location and place it in another particular location. After this process is complete, the subroutine instructs the computer to take its next instruction from the address saved in the special register. Control is then returned to the main program, which includes instructions to take the square root from the second specified location and process it as necessary.

This brief description indicates the main features of the subprogram:

(1) A calling sequence, which stores the data to be operated upon in particular locations defined in both the main and subprograms and stores the address of the next main program instruction in a special register
(2) Modification of the contents of the instruction counter to transfer control of the computer to the subprogram
(3) Execution of the subprogram
(4) A return sequence, in which the results are stored in particular locations defined in both the main and subprograms and modification of the contents of the instruction counter to transfer control of the computer back to the main program

By using the same lengthy sequence of instructions many times, the subprogram makes possible savings both in memory-size requirements and in programing time. Once the subprogram is written, the programer need only write a new calling sequence each time the subprogram is to be used.

CHAPTER 8

Spacecraft Power

GENERAL DISCUSSION

SEVERAL OF THE SUBSYSTEMS of a typical lunar or planetary spacecraft require electrical power for their operation (ref. 10). The major users of electrical power for a typical unmanned planetary probe are the communications, guidance and control, sequencing, and scientific-instrument subsystems. Electrical power is also required by the structural and propulsion subsystems for (1) actuators to move various parts of the structure, (2) solenoid valves, and (3) the firing of pyrotechnic devices (squibs) in explosive bolt cutters, pin pullers, and valves.

A block diagram of a typical spacecraft power supply is shown in figure 8.1. The elements shown correspond roughly to the separate physical elements of the system. Some systems will not contain all of the elements shown; for example, the secondary battery and its associated charger may be omitted for a relatively short flight to the Moon. Of the elements shown—power source, converter, primary and secondary batteries, and the conditioning, control, and charging equipment—the system design is most strongly affected by the selec-

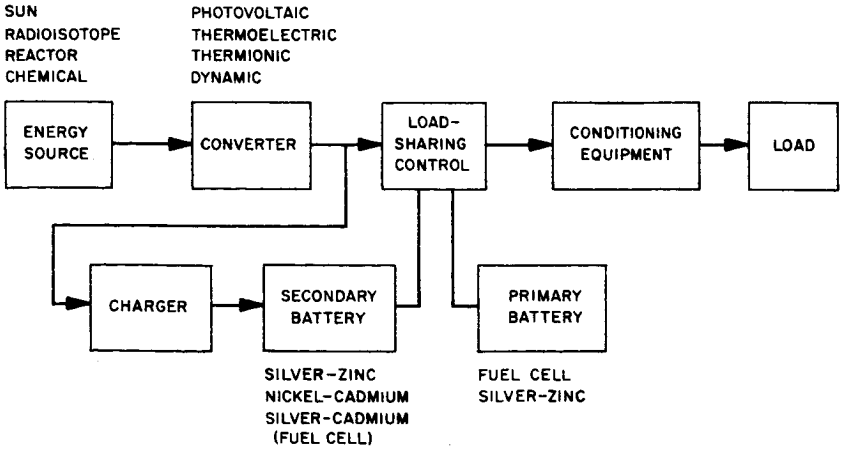


FIGURE 8.1—Block diagram of a typical spacecraft power supply.

238-635 O—67——6

tion of the primary energy source and the converter associated with the primary source.

Of the many types of energy sources and power converters known, only a few are suitable for space missions. Only a subset of these types is suitable for long-life missions, such as a spacecraft to be placed in orbit about Mars at the end of an 8-month flight from the Earth.

For the general range of space missions, practical energy sources include:

(1) Solar radiation
(2) Nuclear reactions
    (*a*) Radioisotope decay
    (*b*) Nuclear fission
(3) Chemical reactions

Practical spacecraft energy and power converters include the following types, which are appropriate for use only with certain sources, as indicated:

(1) Photovoltaic	Solar radiation
(2) Thermoelectric	Nuclear reactions (both types)
(3) Thermionic	Nuclear reactions (both types)
(4) Dynamic	Nuclear reactions (both types)
(5) Batteries	Chemical reactions
(6) Fuel cells	Chemical reactions

Two basic factors determining the applicability of each source-converter combination are:

(1) Power level required
(2) Operating lifetime required

The effects of these factors upon the choice of power-system configuration are shown in figure 8.2. The figure indicates the need for considering nonchemical systems for lunar and planetary missions.

Additional factors affecting the choice of a power system configuration for a particular mission include:

(1) Reliability
(2) Availability
(3) Versatility
(4) Spacecraft interface problems
    (*a*) Temperature control
    (*b*) Attitude control
    (*c*) Radiation fields
    (*d*) Configuration
(5) Weight
(6) Cost

The spacecraft power system must operate to assure even a partial mission completion; thus, reliability is considered to be of paramount importance.

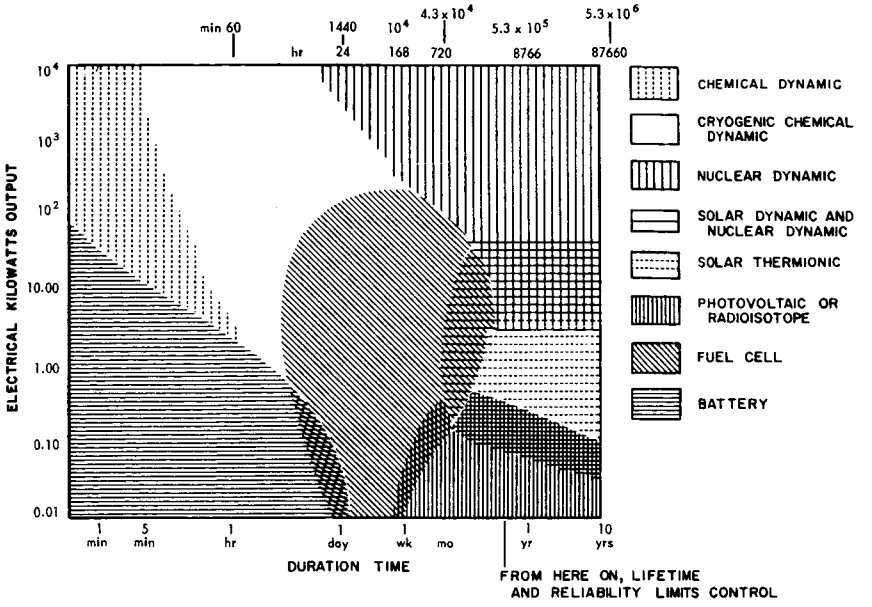


FIGURE 8.2—Estimated 1965 applicabilities of space power systems.

The factor of availability includes two types of problems: compatibility of design, manufacturing, and testing schedules with planetary launching opportunities; and availability of critical components and materials such as radioisotopes, or high-efficiency solar cells, in sufficient quantities to meet the power demands of the missions contemplated.

The remaining characteristics are not necessarily listed in order of importance. The factor of versatility includes the ability of the power system to operate under a variety of mission conditions, e.g., both during the transit from Earth to Mars and in orbit about Mars; and the ability to accommodate variations in the loads and source characteristics. The ability to function at reduced efficiency despite partial failures becomes increasingly important for long missions.

Different power systems present differing interface problems to the remainder of the spacecraft system, some of which have very strong effects upon the system design. Solar-panel size and radioisotope heat are two examples.

ENERGY SOURCE AND CONVERTER CHARACTERISTICS

The characteristics of several combinations of energy sources and converters are described next. Because the same basic conversion principle may be used with different energy sources, the operation of

each converter type is described in detail only once, and cross-references are made under the other possible energy sources.

Solar Energy

Solar energy is received by the spacecraft in the form of electromagnetic radiations impinging on energy-collecting devices. This situation immediately implies the necessity of an unobstructed line of sight from the Sun to the energy-collecting devices, and thus the need for controlling the orientation of these collecting devices.

Such control may be effected by controlling the orientation of the entire spacecraft if the energy collectors are rigidly mounted, or by motion of the collectors with respect to the rest of the spacecraft.

The solar energy is distributed over a spectrum, with nearly all of the energy being radiated in the infrared and visible regions of the spectrum. Because of this distribution, electrical power for the spacecraft can be obtained by either photovoltaic or thermal conversion. In the former method, large-area semiconductor solar cells are used; while in the latter, thermoelectric, thermionic, or dynamic (rotating machinery) devices may be used for conversion.

Solar Photovoltaic Conversion

Of the many semiconductor materials sensitive to solar radiation that could be used to directly convert solar radiation into electrical power, a few, such as silicon, cadmium sulfide, and gallium arsenide, have actually been investigated and used for the manufacture of solar cells. Silicon has received the most attention, and at present is the only material that has been used in flight equipment. Only silicon solar cells will be discussed here.

A solar cell is a P–N junction semiconductor device which converts incident radiant energy in particular portions of the spectrum directly into electrical energy. The characteristics of the solar cell as a source of electrical power depend on the intensity and spectral characteristics of the incident energy, the temperature of the solar cells, and the material and junction characteristics of the semiconductor. An increase in light intensity results in a proportional increase in short-circuit current capability and a logarithmic increase in open-circuit voltage. The open-circuit voltage is approximately inversely proportional to cell temperature.

The characteristics of a solar cell are shown in figure 8.3. A typical cell is 1 by 2 centimeters in size and capable of about 23-mW maximum output, an open-circuit voltage of 0.53 V, and a short-circuit current of 66 mA in space near the Earth. The solar energy inter-

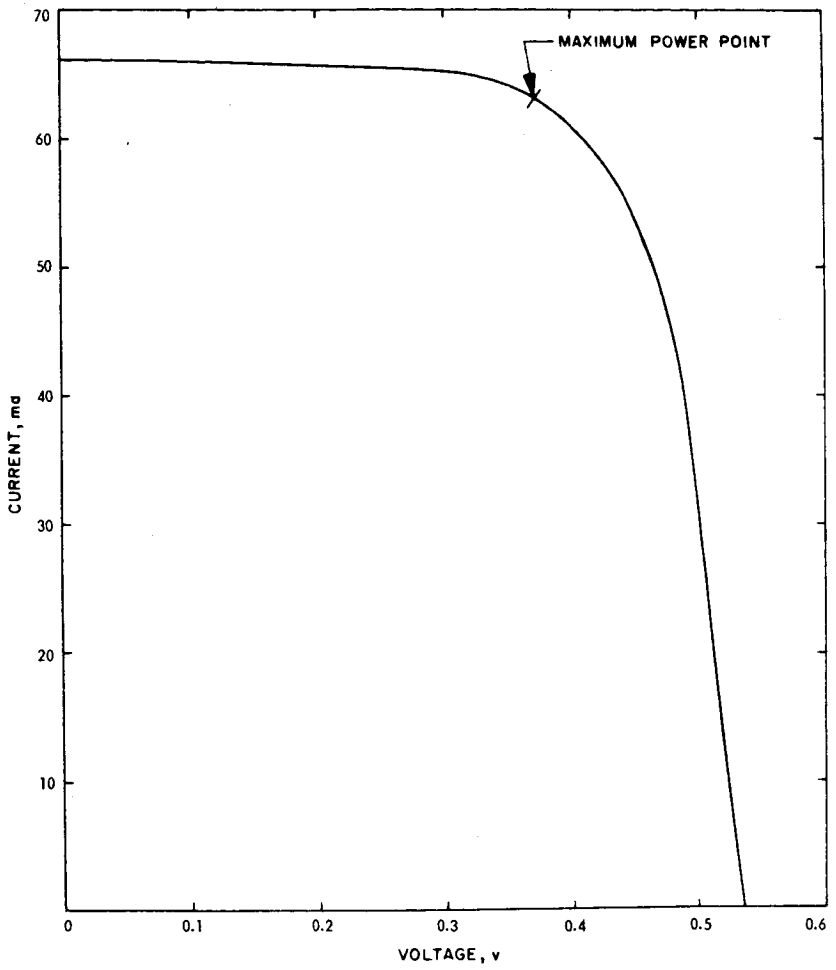


FIGURE 8.3—Solar-cell characteristics.

cepted by the cell, and consequently the electrical power output of the cell, is proportional to the cosine of the angle between the perpendicular to the active surface of the cell and the line from the Sun to the cell. Hence, for maximum power from a given weight of cells, the cells should all be perpendicular to the line from the Sun to the cells. Variations of a few degrees from perpendicularity do not markedly affect the cell outputs.

In a typical lunar or planetary spacecraft, large numbers of cells are laid out on flat panels that are oriented with their faces as nearly perpendicular to the panel sunline as practicable. A sufficient number of cells are connected in series to provide a convenient output

voltage, and blocks of these series strings are connected in parallel. Each series-parallel block is then isolated from the other blocks to prevent an accidental short circuit within one block from short circuiting the entire power supply.

The panel size is then chosen to provide sufficient extra capacity so that the critical spacecraft loads are supplied even when one block is lost. These connections are illustrated in figure 8.4.

A solar panel designed for operation at or near the Earth's distance from the Sun would have a surface area of about 0.1 ft^2/W of power output and a weight of about 1 lb/ft^2 of area. Lighter weight systems have been proposed and built, but at the present they are not satisfactory for surviving a high-acceleration launch environment.

Because the solar-cell output characteristics vary both with temperature and solar intensity, the power capability of a solar panel will vary considerably over the range of space environments from Venus to Mars. Figure 8.5 shows the solar intensity as a function of Sun-probe distance. The variation in solar intensity from Venus perihelion to Mars aphelion is greater than five to one. The temperature of the solar-cell panels varies as a result, as shown in figure 8.6. At Venus perihelion the temperature of a typical panel would be 122° C, and at Mars aphelion, −22° C. The relative performance, affected by both intensity and temperature, is shown for conditions at Venus, Earth, and Mars in table 7.7. The highest output is obtained at Venus, and is approximately three times that available at Mars. Both the solar-panel power capability and the panel voltage vary between Mars and Venus missions.

Solar Thermoelectric Conversion

The basic principles of thermoelectric conversion are described below under "Radioisotope Thermoelectric Conversion."

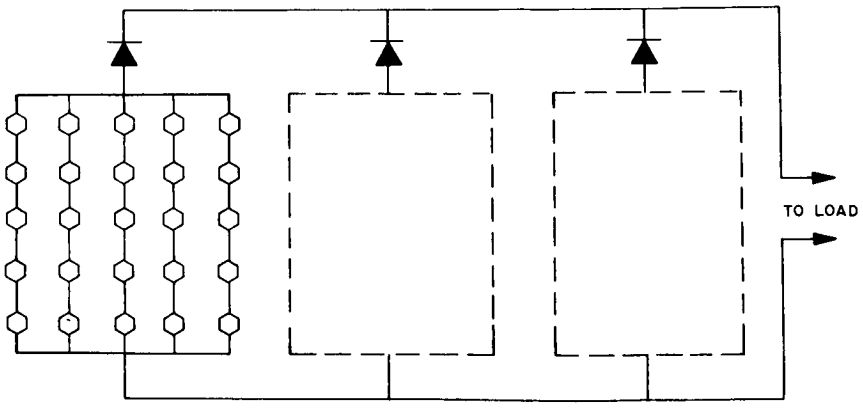


FIGURE 8.4—Typical series-parallel panel.

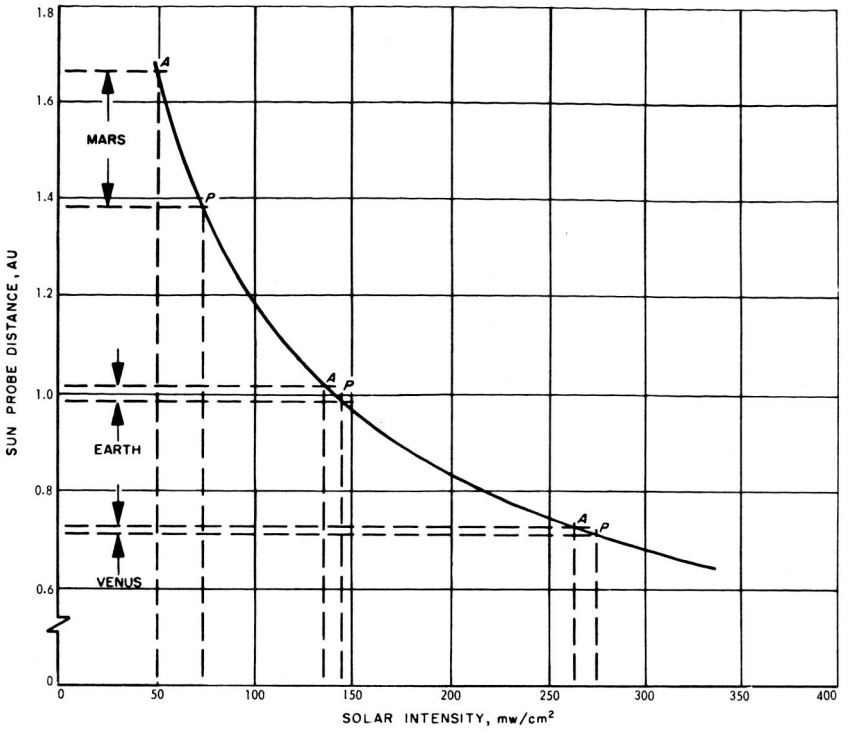


FIGURE 8.5—Solar intensity as a function of Sun-probe distance.

TABLE 7.7—Relative Panel Performance

Parameter	Earth space	Venus		Mars	
		Aphelion	Perihelion	Aphelion	Perihelion
Solar intensity, mW/cm²	140	263	272	50	73
Cell temperature, °C	57	117	122	−22	4
Relative cell efficiency	1.0	0.71	0.69	1.57	1.55
Relative power	1.0	1.33	1.34	0.41	0.59

A solar thermoelectric power supply is heated by solar energy and converts the heat energy directly into electricity. The performance of such devices is dependent on the ability of the collector to absorb and retain solar energy, and on the ability of the thermoelectric elements to convert the resulting heat into electricity. Two different designs are under development for the Air Force Aeronautical System Division. One utilizes a flat-plate collector with the thermo-

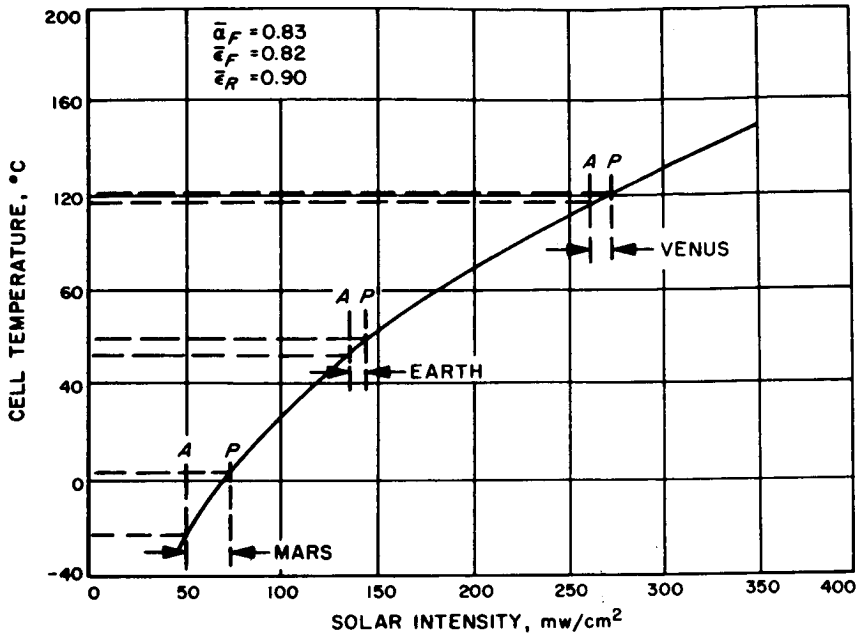


FIGURE 8.6—Variations of cell temperature with solar intensity.

electric elements mounted in a sandwich construction. An alternate design uses a small parabolic concentrator to heat the hot junction of the thermoelement.

The flat-plate design utilizes an absorber surface with a spectrally selective coating. The coating must have a high absorptance in the solar spectrum with a low emittance in the infrared spectrum. The rear surface, in direct contact with the cold junction, must have a high emittance for long-wavelength infrared radiation. The concentrator design requires accurate Sun orientation and surface geometry with a high specular reflectance.

Solar Energy Thermionic (SET) Conversion

Solar energy thermionic electrical power supplies have been under development since 1961. Two configurations of a 500-W SET power source were studied for possible use on the early designs of the Mariner-Mars spacecraft. However, the development of SET power supplies had not progressed sufficiently by 1964 to warrant selection for the 1964 Mariner-Mars mission.

The thermionic conversion process is shown schematically in figure 8.7. Solar energy is collected with a parabolic mirror, reflected and concentrated into an intense beam and used to heat the emitter

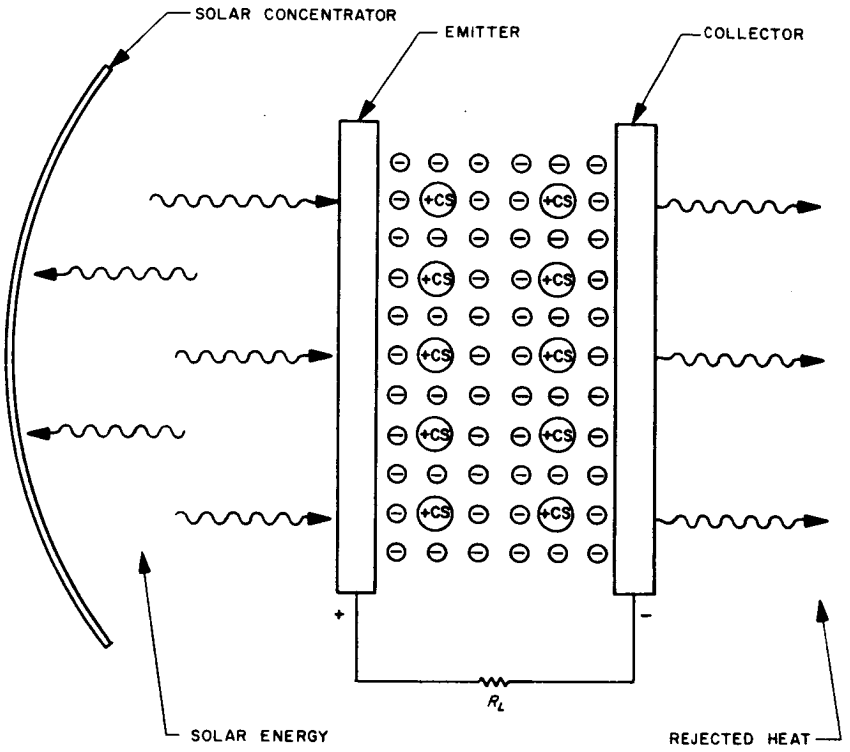


FIGURE 8.7—Schematic of thermionic energy conversion process.

of a thermionic converter to a temperature between 1500° and 2500° K. Electrons are evaporated from the hot emitter surface and travel through the interelectrode space to the collector. The ionized cesium atoms neutralize the space charge that would otherwise exist as the result of the large electron current. The excess electrons at the collector travel through the external load and return to the emitter, completing the circuit. The solar energy not converted to electricity is transferred to the collector and is radiated as heat. The output voltage of the converter is the difference between the emitter and collector Fermi levels.

A cross section through a typical SET converter is shown in figure 8.8. Solar energy impinges upon the surface of the emitter and cavity piece. A fraction of the incident energy is absorbed and heats the emitter to approximately 2000° K. The space between the emitter and collector is approximately 0.002 inch at operating temperature. This space is filled with partially ionized cesium gas at a pressure between 2 and 6 mm Hg. The hot emitter has a partial coverage of adsorbed cesium which lowers the emitter work function and increases

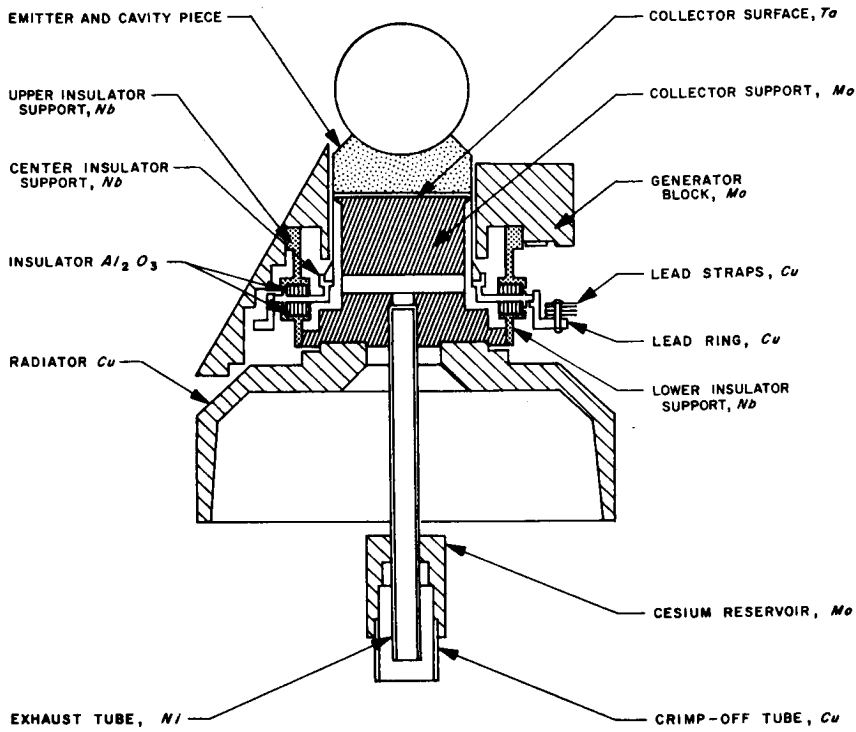


FIGURE 8.8—Cross section of SET converter.

electron emission. The collector is covered with at least a monolayer of adsorbed cesium which lowers the collector work function to a value between 1.7 and 1.9 electron volts. The converter is attached to a molybdenum generator housing through an insulated niobium support ring. The emitter and collector are electrically insulated by alumina joined to niobium sections. The molybdenum collector body is attached to a copper radiator coated with chromium oxide. The cesium reservoir is located at the end of a nickel exhaust tube, and is heated to approximately 360° C by conduction and radiation from the collector and radiator.

Typical output voltages are in the 1-V range; and conversion efficiencies of up to 15 to possibly 20 percent are expected as converter designs and techniques are refined.

Solar Energy Dynamic Conversion

In dynamic conversion of solar energy into electricity, solar energy is used to heat a working fluid, such as mercury, which is vaporized and then expanded through a turbine to drive an alternator. In the

Sunflower system, a parabolic reflector is used to focus solar energy on a mercury boiler. The rotating machinery is similar in concept to that being developed for the nuclear-reactor dynamic power systems.

Nuclear Energy From Radioisotope Decay

A radioisotope electrical power generator derives its energy from the radioactive decay of an isotope. Unstable isotopes, obtained either as byproducts of a fission reactor or formed by neutron bombardment in a breeder reactor, decay by emitting particles or electromagnetic radiation containing large amounts of energy. The types of events that occur during this decay include (1) emission of the alpha particles (helium nuclei), (2) emission of beta particles (electrons originating from the nucleus), (3) emission of electromagnetic quanta called gamma radiation, (4) spontaneous fission (which occurs only in very large nuclei), yielding two smaller nuclei plus gamma radiation. The energy liberated during the decay process is transformed into heat. By proper choice of radioisotope, containment materials, and physical arrangement, a thermal source of high heat density per unit of volume and weight can be constructed. This thermal source is combined with a suitable conversion device to give a self-contained radioisotope power generator.

The radioisotope fuel selected for a power system should—

(1) Have a sufficient power density to match the requirements of the energy converter;
(2) Have a sufficiently long half-life[1] to give a relatively flat power output for the operating-time requirement of the system;
(3) Be available in sufficient quantity at the required time to be used for a given flight program; and
(4) Have a limited amount of hard-to-control, high-energy radiation.

The first criterion, power density, will be determined by the type of energy converter used. When a thermoelectric type of energy converter is considered, a lower temperature at the hot junction will

[1] The thermal power output decreases with time according to:

$$P(t) = P(t_0)e^{-K(t-t_0)}$$

where

t	time
t_0	an initial time
P	power
K	a positive constant
e	base of natural logarithms

The *half-line* T is that value of $t-t_0$ for which

$$-KT=\ln (0.5)$$

so that

$$P(t_0+T)=0.5\ P(t_0)$$

mean a lower conversion efficiency, although power can be obtained from any isotope. Because thermionic converters require high temperatures, only certain fuels are suitable for use with them.

The second requirement, long half-life, is dictated by the mission. For a 1-year operating time, half-lives on the order of years become very desirable. This requirement may be offset, however, by the use of power output flattening devices (such as a highly damped, thermal-hydraulic shutter control system which controls radiation of the time-predictable excess heat of the radioisotope decay process). However, devices of this type tend to become complicated and detract from the inherent simplicity and reliability of the radioisotope power supply. Figure 8.9 illustrates the effect of the half-life of several radioisotopes on normalized output power as a function of time.

The third requirement, availability, may be by far the most important, since at present none of the radioisotopes desirable from the point of view of the other criteria is available in sufficient quantity to fuel a very large spacecraft power system. This, however, may be remedied by adequate leadtime for the development of manufacturing facilities.

The fourth criterion will be defined by the ground-handling requirements and radiation tolerance of the equipment and experiments in the spacecraft.

For operating periods (including both preflight tests and flights) of 1 year, only three radioisotopes seem to meet the half-life criterion: strontium 90 with a half-life of 28 years, plutonium 238 with a half-life of 89 years, and curium 244 with a half-life of 19 years. The only one of these isotopes that is readily available is strontium 90, a common product of fission reactors; the other two are transuranic elements produced in breeder reactors. Their production at present is very limited.

Strontium 90, however, has two main disadvantages. One is that it has low power density, a maximum of 3.8 thermal W/cm^3, which results in a heavy system of low efficiency. The other is that it presents serious radiation problems. Plutonium 238 and curium 244 are principally alpha emitters. Plutonium 238 produces some low-energy gammas in addition to alphas. In larger systems, this may become serious. Plutonium 238, with a power density of about 9.5 W/cm^3, is suitable for a high-efficiency thermoelectric system. Curium 244 produces pure alphas and has no gammas in its regular decay; however, a very minute fraction of this decay is in the form of "spontaneous fission." This does not become a problem, however, until the amount of fuel used in a system becomes large. Because curium 244 produces 2.7 thermal W/gram, a 5-percent-efficient

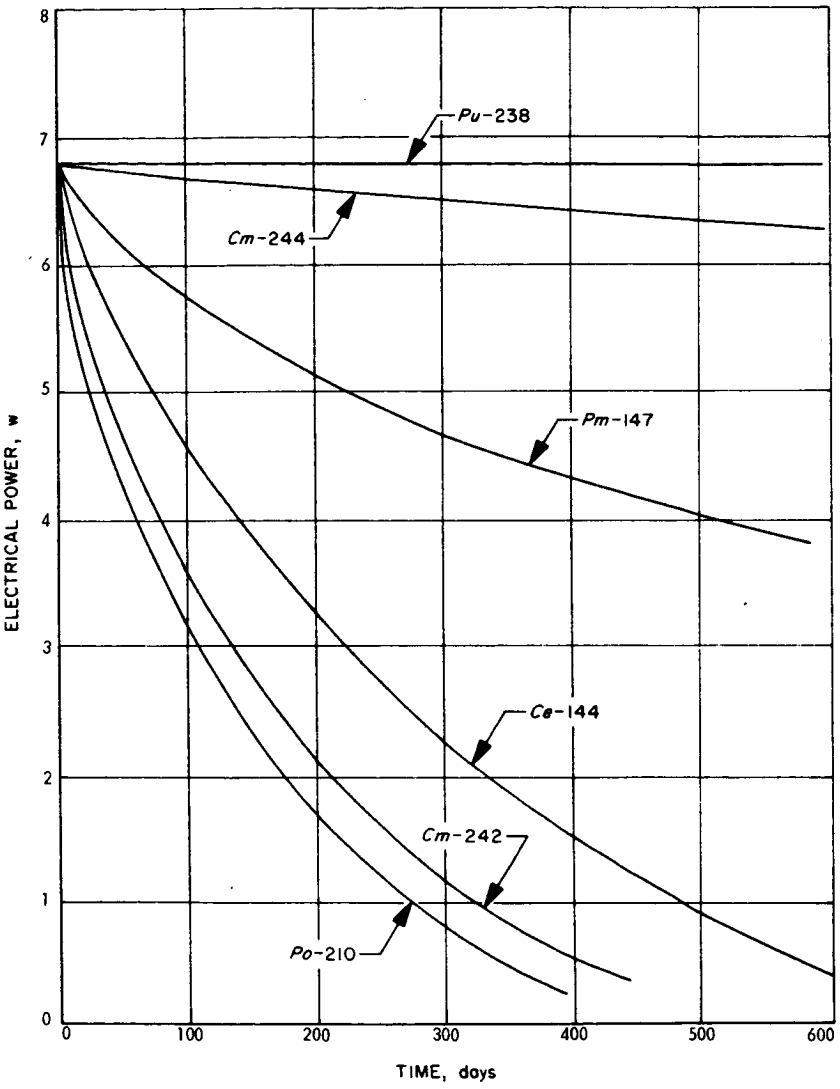


FIGURE 8.9—Isotope electrical power (watts) versus time.

thermoelectric system will require about 7.5 kg/kW electrical. This is sufficient to put a 300-W electrical power system in the "large" class, which would require special handling procedures because of the neutrons produced by the fission.

The problem of availability for both these desirable fuels (Pu^{238} and Cm^{244}) is of about the same magnitude. However, more is currently known about Pu^{238} and its applicability to radioisotope power systems.

Radioisotope Thermoelectric Conversion

Thermoelectric converters use the Seebeck effect to convert heat into electrical energy. Most materials show a migration of electrons from a hot end to a cold end if a temperature gradient is established. The ideal material would combine the properties of a perfect thermal insulator with those of a perfect electrical conductor. No known material meets these requirements, but the best compromise is found in semiconductors. Semiconductors are of two types: *n*-type with an excess of electrons, and *p*-type with a deficit of electrons. When these two types of material are combined to form a thermocouple, the voltage output of the device is improved.

A thermoelectric energy converter is a heat engine and is, therefore, subject to the limitations of the Carnot cycle. Hence, the efficiency is proportional to the temperature difference that can be imposed on an element. To maximize this difference, it is desirable both to increase the hot-junction temperature and decrease the cold-junction temperature. The hot-junction temperature is limited by difficulties with the properties of materials, such as thermoelectric elements, the heat source, structures, conductors, and insulators. The cold-junction temperature is limited by practical radiator sizes, because the only process for the rejection of waste heat in space is radiation. The radiator area is inversely proportional to the fourth power of the desired cold-junction temperature. For example, a radiator-temperature decrease from 350° to 300° C requires an area increase of 50 percent if the same amount of heat is to be rejected. Thus, if materials can be developed to operate at higher temperatures, a more compact, more efficient system can be obtained. However, this fact is tempered by the problem of integrating the power system into the spacecraft. The power system is a source of heat, and thus its location within the spacecraft is critical with respect to temperature-sensitive components.

Radioisotope thermoelectric systems have been under development since 1956, and have resulted in flight systems such as those for the Transit satellites. Lead telluride, the thermoelectric material currently used, is capable of achieving a specific power output of 1 W/lb. An alternate design would replace the *p*-type lead telluride with zinc antimonide, which may yield a higher specific power output. Another design being studied replaces the lead telluride with a germanium-silicon alloy, which is expected to increase the output to about 1.5 W/lb for generator output up to 25 W. Additional improvement may be possible for 50-W units. These performance estimates are predicated on the use of plutonium 238 as the fuel.

Radioisotope Thermionic Conversion

For some fuels, sufficient heat can be obtained to replace the thermoelectric converter with a thermionic converter similar in principle to that discussed under "Solar Energy Thermionic Conversion."

The description of the thermionic conversion process given under solar thermionic conversion is applicable here, except that the solar heat source is replaced with a radioisotope heat source and the emitter temperature is somewhat lower. Fuels with high reactivity are required to obtain the required operating temperature. This eliminates the low-power-density, long-half-life beta emitters. The fuels considered are cesium 144, curium 242, polonium 210, and possibly plutonium 238. The first three, due to their shorter half-life, require a mechanism for flattening the thermal power delivered to the conversion device. The use of cesium 144 may also result in serious handling difficulties because of the radiation emitted. Most of the pure elements have melting points which are below the required operating temperature of the thermionic generator. It is necessary to use isotopes in fuel forms such as carbides or oxides to obtain a high melting point. The dilution of the isotope reduces the thermal power density (W/gram) that can be obtained.

Radioisotope thermionic generators are expected to be more efficient than thermoelectric generators and may be capable of a power output of 5 W/lb. Generators in the 100- to 500-W power level have been under study for some time.

Nuclear Energy From Fission Reactors

Three SNAP systems for nuclear auxiliary power reactor systems of possible application to advanced spacecraft are under development by the Atomic Energy Commission and the National Aeronautics and Space Administration.

A typical reactor for thermoelectric conversion would be a thermal reactor using zirconium hydride as a moderator and uranium 235 as the fuel. Beryllium could be used as a neutron reflector around the reactor core. Radiological safety requires that the reactor must not be started until the spacecraft has successfully been injected into orbit.

A typical reactor coolant is a eutectic mixture of sodium-potassium (NaK). The coolant could be circulated using an electromagnetic pump operated by an independent bank of thermoelectric elements. The coolant loop can also provide vernier control to reactor power level because an increase in the reactor fuel rod temperature raises the reactor coolant temperature, causing an increase in the power output

of the thermoelectric bank operating the coolant pump. The coolant is pumped at a higher rate, thereby decreasing the reactor fuel rod temperature. Reverse operation can also occur, since a decrease in reactor fuel rod temperature will decrease the coolant flow rate, causing the reactor fuel rod temperature to increase.

The thermoelectric generator might consist of a bank composed of several strings of thermoelectric elements. The hot junctions of the thermoelements could be mounted to the coolant pipes. Each thermoelectric couple would then have a strip of metal to form the cold junction. The metal strips together form the waste-heat radiator.

An early design of such a system utilized lead telluride thermoelectric elements. The system temperature was limited to 900° F at the hot junction by the properties of lead telluride; above this temperature this material appears to be marginal for an operating life of 1 year. A germanium-silicon-alloy thermoelectric element suitable for integration into the system has been developed to increase the hot-junction temperature.

Other reactors are being developed to use the same rotating machinery as the Sunflower solar dynamic system. Such a system might be used to provide electrical power for low-thrust ion propulsion of an interplanetary spacecraft.

Chemical Energy Conversion With Batteries and Fuel Cells

Batteries and fuel cells are similar in that both convert chemical energy directly into electrical energy without an intervening heat engine or conversion machinery (such as turbines). Both devices can be obtained in either primary configurations, which use expendable reactants and cannot be recharged, or secondary configurations, in which the reactants can be reused after chemical energy has been restored to them. The secondary cells permit the use of charge-discharge cycles for carrying intermittent loads or supplying power during periods when the primary power source is inoperative; e.g., a solar-powered system when the spacecraft is in the shadow of the planet.

The major differences between fuel cells and batteries are in configuration, development status, specific energy (W-hr/lb), and operational problems.

The battery is essentially an inseparable combination of energy source and converter, where the reactants are contained within the cell. Batteries are well-developed items, with specific energies of primary cells in the range of 10 to 100 W-hr/lb and of secondary cells in the range of 2 to 40 W-hr/lb. Operational problems of bat-

teries include a restricted operating-temperature range (0° to 150° F), relatively short shelf life after activation for certain types of primary cells, and the necessity of keeping certain configurations nearly upright during ground and launch operations. Three battery types that may be considered for space missions are the silver oxide-zinc cell, the nickel-cadmium cell, and the silver-cadmium cell. The silver oxide-zinc cell may be used either as a primary battery or as a secondary battery for a limited number of charge-discharge cycles. This type of cell has a specific energy of 40 W-hr/lb and a specific volume of 2.7 W-hr/in^3. For application requiring many charge-discharge cycles, the sealed nickel-cadmium cell may be used. This cell has a specific energy of 10 W-hr/lb and a specific volume of 1.0 W-hr/in.^{3} The silver-cadmium cell is currently being investigated; its properties represent a compromise between those of the other two types mentioned here.

In the fuel cell, a stream of reactants from separate tanks is passed through a conversion device. Fuel cells for manned space applications have been developed. For some phases of unmanned planetary spacecraft missions, the hydrogen-oxygen primary fuel cell may be considered. This device has a specific energy of 150 to 300 W-hr/lb, depending upon the storage method used. A fixed weight for the conversion cell and auxiliary equipment must also be added. Operational problems include control of the reaction, long-term storage of the reactants either as cryogenic liquids or as high-pressure gases, and the lack of experience in applications of fuel cells to unmanned space missions. Fuel cells are not recommended as the primary power source for a long unmanned mission; however, they may be used to supply peak loads in a solar- or radioisotope-powered system.

DESIGN FACTORS

It was stated at the beginning of this chapter that the important factors affecting the choice of the energy source and converter were:

(1) Power level required
(2) Operating lifetime required
(3) Reliability
(4) Availability
(5) Versatility
(6) Spacecraft interface problems:
    (*a*) Temperature control
    (*b*) Attitude control
    (*c*) Radiation fields
    (*d*) Configuration
(7) Weight
(8) Cost

238-635 O—67——7

Some brief remarks of the effects of the design factors upon the choice of power system energy source and converter are listed below:

(1) Dynamic (moving machinery) conversion systems do not currently appear to have the reliability potential for use in a 1-year unattended space mission.

(2) Storage of chemical reactants as high-pressure gases or as cryogenic liquids introduces a number of problems in the design of the spacecraft system. Both storage weight and storage lifetime must be considered.

(3) Nuclear reactors are relatively heavy, and produce high radiation fields, which would require heavy shielding to prevent the accumulation of excessive radiation doses by other spacecraft components during a 1-year operating lifetime. The reactor also must be of at least a minimum size.

The remaining systems are solar or radioisotope powered and use static converters; they include photovoltaic (solar only), thermoelectric, and thermionic systems, outlined as follows:

(1) Solar photovoltaic system—

(*a*) Has interface requirements, such as requirements for solar orientation, deployment and adequate rigidity of large-area solar panels, and effects upon the heat balance of the spacecraft.

(*b*) Is unable to operate in the absence of solar radiation, e.g., during maneuvers and in the shadow of a planet; this factor also affects preflight ground checkout and systems tests, and requires a large energy-storage unit (battery or fuel cell).

(*c*) Has solar cells susceptible to degradation in performance from solar flare and Van Allen-belt types of radiation.

(2) Solar thermionic system—

(*a*) Has more stringent interface requirements on solar orientation and deployment mechanisms.

(*b*) Is unable to operate for long periods in absence of solar radiation; however, such a thermal system can be provided with a short-term heat-storage element which may increase its reliability and reduce the size of the electrical storage system.

Two advantages of this system are its promising high conversion efficiency and light weight. A disadvantage is its less-advanced state of development.

(3) Solar thermoelectric system—

(*a*) Has effects on other spacecraft systems; these effects are similar in magnitude to the photovoltaic system for the flat-plate type, or are more stringent (comparable to the thermionic system) for the concentrator type. The sup-

ports and deployment mechanisms still must be developed and may reduce or eliminate some of the weight advantages these systems appear to have.

(*b*) Is unable to perform in the absence of solar radiation; these systems are less amenable to a heat storage system than the solar thermionic system.

(4) Radioisotope thermoelectric system—

(*a*) Has the principal problem of the availability of radioisotope fuel; this problem may be very critical to this otherwise very desirable system.

(*b*) Has the effects of heat and nuclear radiation upon the performance or design of other spacecraft components.

(5) Radioisotope thermionic system—

Has the problems of the radioisotope thermoelectric system; in addition, the development work required may take several years. However, the system has promise of higher conversion efficiency which will require less fuel and may enable reduction of system specific weights.

All of the systems described above suffer from the inability to handle short-duration loads above the steady-load capacity of the system. Consequently, all systems require a battery or fuel-cell energy-storage unit to handle peak demands of short duration.

POWER-CONDITIONING EQUIPMENT

The loads present in the spacecraft may present widely differing demands in terms of voltage level, voltage regulation, power demand, load and switching profile, frequency stability, etc. In addition, the output characteristics of the energy source and converter may vary widely with load and with operating conditions.

Broadly stated, it is the task of the power-conditioning equipment to transform the converter output into forms suitable for consumption by the various loads. To accomplish this task, the conditioning equipment must perform the functions of voltage regulation, frequency control, and distribution.

In small, highly integrated, special-purpose spacecraft with only a few loads, it may be convenient to transform the converter output directly into the forms required by the individual loads. However, as the spacecraft complexity and number of individual loads increase, it becomes advantageous from a system-design standpoint to introduce an additional step into the conditioning process. The output of the converter is transformed to a small number of standard distribution voltages.

The voltage and frequency of these standard voltages are regulated by the central power subsystem of the spacecraft in such a way that

the majority of loads can be supplied by further simple transformation at the load end.

Some of the types of loads encountered on a typical planetary spacecraft include such diverse types as—

(1) High-current pulse loads, such as squibs and solenoids
(2) Electronic equipment, requiring several dc voltages with reasonably good regulation
(3) High-current, high-voltage loads, such as the power required by the final stage of the transmitter
(4) High-voltage, low-current loads, for deflection circuits in the vidicons and image dissectors used in data gathering and optical guidance and control sensors
(5) Ac actuator motors not critical with respect to line frequency or voltage
(6) Ac motors requiring extremely accurate frequency control.

The block diagram of a typical power system is shown in figure 8.10.

The load-sharing control is necessary to avoid draining the battery when the solar panels are capable of carrying the entire load. The solar panels and battery are diode isolated from each other to keep from draining battery power into the solar panels when they are not capable of carrying the entire load.

As discussed under "Solar Photovoltaic Conversion," the output voltages of the solar panels vary widely as the load demanded from the panels varies. It is the function of the regulator to absorb these variations. There are a number of practical regulator configurations based on either of two fundamental principles. In one method, the voltage drop across a resistor in series with the load is varied to maintain the voltage drop across the load constant. This type of regulator thus dissipates power as heat in the series resistor.

In the other type of regulator, losses are minimized by withdrawing power from the source in pulses and passing these pulses into an inductance-capacitance filter which smooths the variations in output voltage. This type of regulator is shown in figure 8.11.

As the block diagram shows, the switching regulator is a servo system. The dc output voltage is compared with the reference voltage, and the difference, or error, is used to control the width of a constant-frequency pulse train that controls the transistor switch. This switch alternately connects and disconnects the transformer and the converter output. As a result, an ac wave is obtained at the secondary of the transformer. This wave is rectified, and fed to the filter. The average dc value of this rectified wave is obtained at the output of the filter. This average value is a function of the fraction of the cycle during which the transformer is connected to the source. The input filter capacitor is used to smooth the demands on the source; it charges during the "off" periods of the switch and discharges during

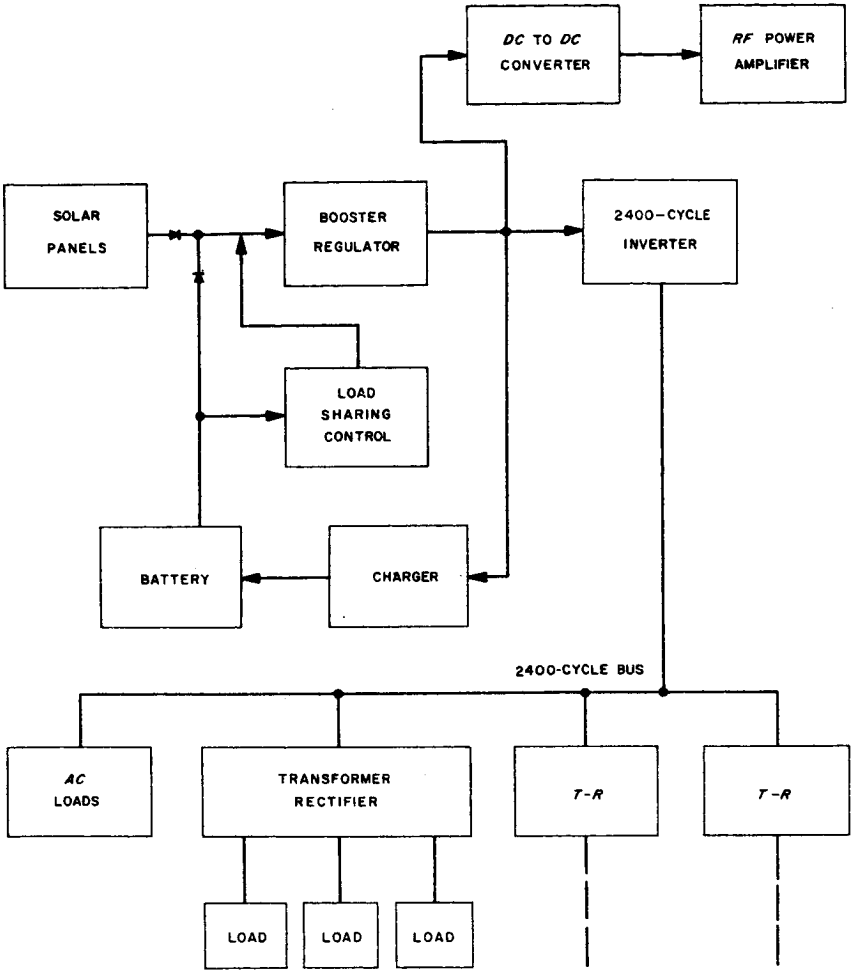


FIGURE 8.10—Block diagram of typical power system.

the "on" periods, thus making the drain on the converter very nearly constant.

The inclusion of the transformer isolates the dc output from the converter. This permits connection of the switching-regulator output in series with the converter output, as shown in figure 8.12, to form a booster regulator, which is capable of regulating the dc output as long as the maximum voltage of the converter is less than the regulated dc output desired, but greater than a lower limit that is a function of the regulator parameters.

The remaining power-conditioning devices operate from the regulated dc output of the booster regulator, and use elements similar to portions of the booster regulator.

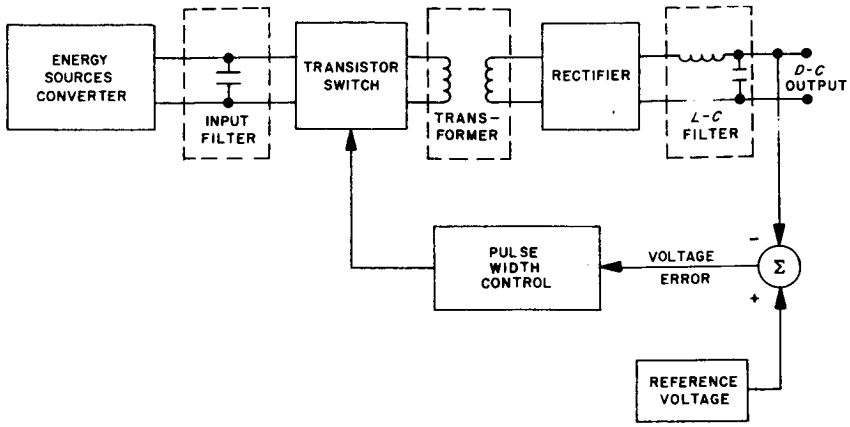


FIGURE 8.11—Switching regulator.

The 2400-cps inverter is basically a transistor switch and transformer with a symmetrical square-wave output. If it is necessary to control the frequency accurately, the switch may be synchronized by an accurate frequency source such as a crystal oscillator followed by a divider chain.

The main power distribution of the spacecraft described here is by the ac 2400-cps square-wave voltage derived from the inverter. This standard distribution voltage is converted to the various dc voltages required by a particular subsystem by a transformer-rectifier which is basically a transformer connected to the ac line, followed by rectifiers and filters.

The RF power amplifier represents a large load at a single high dc voltage. The conversion of the dc voltage from the booster regulator level to the level required by the power amplifier is accomplished with a dc-to-dc converter, which is a transistor switch followed by a transformer, rectifier, and a filter, in an arrangement similar to that used in the switching regulator.

The terminal voltage of the battery rises as the charge stored in the battery increases. Because overcharging the battery will damage it, the battery charger must stop charging when the terminal voltage rises above a predetermined level.

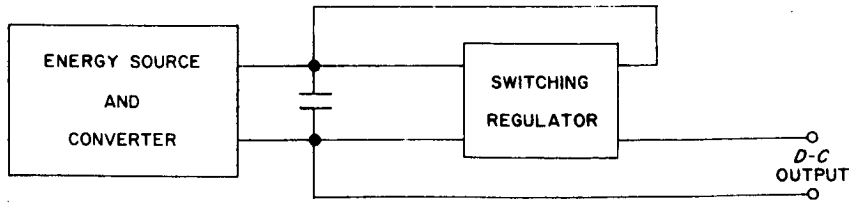


FIGURE 8.12—Booster regulator.

CHAPTER 9

Spacecraft Control Systems

SPACECRAFT CONTROL SYSTEMS are required to perform several functions. One that is often thought of first is attitude control. Attitude control is defined as control of the rotational motion or orientation of the spacecraft about its center of mass. Also, this term is usually used in connection with the unpowered or coasting phase of a mission. During powered portions of a mission, an autopilot is used to maintain the proper orientation of the space vehicle. A final function performed by spacecraft control systems is that of moving or pointing articulating members such as antennas, sensors, or cameras in the proper direction.

ATTITUDE CONTROL

The need for attitude control of a space vehicle arises from several sources. Solar cells mounted on solar panels must be pointed toward the Sun. From a temperature-control standpoint, it is desirable to have the Sun in a fixed position relative to the spacecraft. Both of these requirements necessitate attitude measurement and control to only moderate accuracy (1° to 5°). Other orientation requirements may derive from the scientific instruments, which may necessitate alinement toward the Sun or a planet. The most exacting accuracy demands, however, come from guidance requirements. To perform trajectory corrections, it is necessary to command the vehicle to an arbitrary orientation to aline the thrust vector of the spacecraft rocket motor to possibly ¼° or better.

The choice of reference system to be used for attitude-control purposes is relatively straightforward for the transit phase of a mission; i.e., the period from injection near the Earth to planetary approach. Since the requirements for this phase involve simultaneous pointing of an antenna at the Earth and solar panels at the Sun, four angular degrees of freedom are required. Two of these degrees of freedom could be motion about the pitch-and-yaw axis of a spacecraft to point the roll axis toward the Sun. The other two degrees of freedom could be motion about the roll axis of the spacecraft and motion of the antenna about a hinge attached to the vehicle. De-

pending upon the mission, the Sun and Earth would be sufficient as reference bodies for the attitude-control system. This was the reference system used for the Mariner II spacecraft that performed a Venus flyby mission.

For geometrical and physical reasons, however, the Earth is often a poor reference body for planetary missions. This is particularly true for trajectories away from the Sun, such as those to Mars or Jupiter. Its luminosity is low because it is seen as a crescent. It is difficult to track because, during part of all trajectories away from the Sun, the Sun is seen by an Earth sensor, approximately 12 orders of magnitude brighter. Thus, for trajectories away from the Sun, some other celestial object, such as the star Canopus, is preferable.

Canopus is desirable as a second-attitude reference because it is situated about 15° from the south ecliptic pole. Since it is the second brightest star, its brightness also makes it distinct from other conspicuous stars. Star tracking requires two additional degrees of freedom, but the relatively small range of Sun-spacecraft-Canopus angles greatly simplifies the pointing problem.

If the attitude-control system is being designed for a planetary orbiter, the nature of the orientation problem changes, and the choice of reference system is no longer as straightforward. If solar power is employed, the requirement for pointing the solar cells is still imposed when the Sun is visible, and high-gain antenna pointing is still required when the Earth is not occulted by the planet. In addition, the planet is the only possible reference body that is never occulted.

It is not at all obvious what the optimum reference system is for this case. If the Sun is not used as a source of power, the choice becomes even more difficult because there are fewer reasons for using it as a reference. Ultimately each mission must be considered individually in order to trade off the advantages and disadvantages of Sun versus planet orientation.

There are several different classes of actuators which are used in attitude-control systems. Typical torques required of these actuators for a Ranger or Mariner class of spacecraft are 0.01 to 0.03 ft-lb. The first and probably most common class of attitude-control actuator includes the mass-expulsion devices. The working fuel is stored on the spacecraft, and a small quantity is expelled from the spacecraft in the proper direction whenever it is desired to apply a torque to the vehicle. The fuel for these devices may be stored gas, liquid which is decomposed into a gas or vapor, liquid which is burned directly in a combustion chamber of a thrustor, or a subliming or decomposing solid. Because all the fuel must be carried along on the spacecraft, all mass-expulsion actuation systems have a limited lifetime.

A second class of actuators is that of the momentum exchange devices. To obtain a control torque, the spacecraft torques against a reaction wheel or sphere; hence, momentum from the spacecraft is transferred to the reaction element. The net angular momentum of the spacecraft system, including the reaction device, does not change during this operation. Momentum interchange devices have been successfully used in satellites where disturbing torques are cyclical; however, for unidirectional disturbance torques, the speed of the reaction wheel or sphere will continue to increase to dangerous speeds. In this case some other means must be used to "desaturate" the reaction element. Because disturbance torques on interplanetary spacecraft tend to be unidirectional, momentum interchange devices have not been used. For the case of cyclical disturbances, the life of these actuators is limited only by the lifetime of the driving electronics or reaction wheel bearings.

A final class of actuators is the field-effect devices. Included are those which interact with a magnetic, gravitational, or solar radiation field to produce useful control torques. These actuators are somewhat limited in their application because they are capable of producing only very small torques; for example, the typical maximum torques produced by a solar-pressure control system on a Mariner type of spacecraft are 100–200 dyne-cm.

As an example of an attitude-control system, consider that of the Mariner IV spacecraft which flew near to Mars in July 1965. The roll axis of the vehicle was pointed at the Sun by means of the pitch-and-yaw control systems, and the spacecraft position about the roll axis was determined by a star sensor pointed toward Canopus. A block diagram of the pitch or yaw axis control system is shown in figure 9.1.

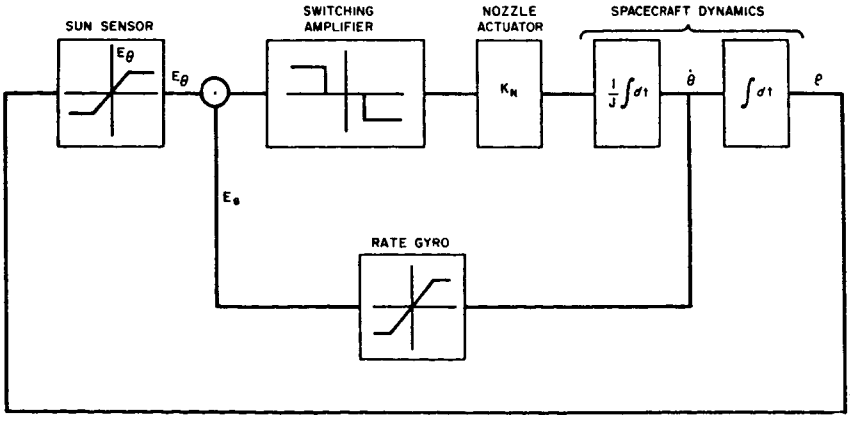


FIGURE 9.1—Pitch-or-yaw attitude control loop.

After the spacecraft is separated from the boost vehicle, the angular rates about each of the body axes are sensed by gyros and are reduced by firing cold-gas (nitrogen) thrustors in the proper direction. The saturation voltages of the Sun sensors and the gyros are chosen so that for rates above a certain value, the gyro signal is greater than the maximum Sun sensor output. When the body rates are reduced below this value, the signal from the Sun sensors has an effect and is used to orient the spacecraft toward the Sun. The function of the gyros then is to provide damping. It can be seen from the switching amplifier mechanization that the thrustors are operated in an on-off manner. This is done primarily to simplify the mechanization and to reduce valve leakage. When the Sun is acquired in both the pitch-and-yaw axes, the spacecraft rolls about the sunline until the star tracker locates the star Canopus, at which time the roll is stopped. The spacecraft has then acquired its references.

There are many more facets of the attitude-control problem that could be discussed. For example, the combined use of two different classes of actuators might offer some advantages in certain instances, or tradeoff studies of the choice of reference system could be mentioned to show the influence on attitude control system design. Also, the effects of spacecraft size or weight could be discussed at some length. The example given, however, should serve to illustrate a typical application of the principles involved.

AUTOPILOTS

After the spacecraft has been injected into its trajectory and the attitude-control system has acquired its references, the vehicle is in a cruise mode until any necessary trajectory corrections are determined. To make corrections, the spacecraft carries a rocket motor capable of adding a small increment to the spacecraft velocity. To accomplish this velocity change, the spacecraft is reoriented to point the thrust vector in the proper direction and the motor is fired for a length of time that gives the desired increment. During this motor-burning phase, an autopilot is used for control.

The purpose of an autopilot is to keep the thrust vector of the rocket motor oriented in the proper direction during the powered flight portions of a mission. In most spacecraft applications it is desired to maintain a constant thrusting direction, although there are a few applications where this is not the case.

Inertial references are usually used in autopilot control systems. Gyros are used to indicate angular motion about each of the three spacecraft axes, and accelerometers provide information on translation of the space vehicle. In some instances, no accelerometers are used.

The velocity change along the thrust direction can then be controlled by calibrating the thrust level of the motor and firing it for a predetermined length of time. Also, side velocities can be estimated by knowing the time history of the angular motion about the body axes and the magnitude of the thrust vector.

In the autopilot control system the required torques are much larger than those required for attitude control. Typical control torques for a Ranger or Mariner type of spacecraft range up to several foot-pounds. There are several techniques that can be used to obtain these torques. On the Ranger and Mariner, jet vanes are used. These are vanes placed in the exhaust of the rocket motor and are used to deflect part of the rocket exhaust to obtain the control force. Other thrust-vector-control techniques gimbal the entire motor or utilize secondary injection. When secondary injection is used, a fluid is injected somewhere downstream of the nozzle throat in order to deflect the thrust vector. These and other thrust-vectoring systems are extremely dependent on the requirements of the mission, and tradeoff studies are required to select the most suitable approach.

A simplified block diagram of an autopilot control loop for a single axis is shown in figure 9.2. The spacecraft dynamics shown are for a single pure inertia, when in actuality the spacecraft is usually represented by several masses connected by springs. The spacecraft angular position and rate are sensed by a gyro with the appropriate electronics. These two signals are combined to form an error signal which drives an amplifier containing any compensation networks necessary to give the desired performance. Finally, this amplifier controls the actuator which torques the spacecraft.

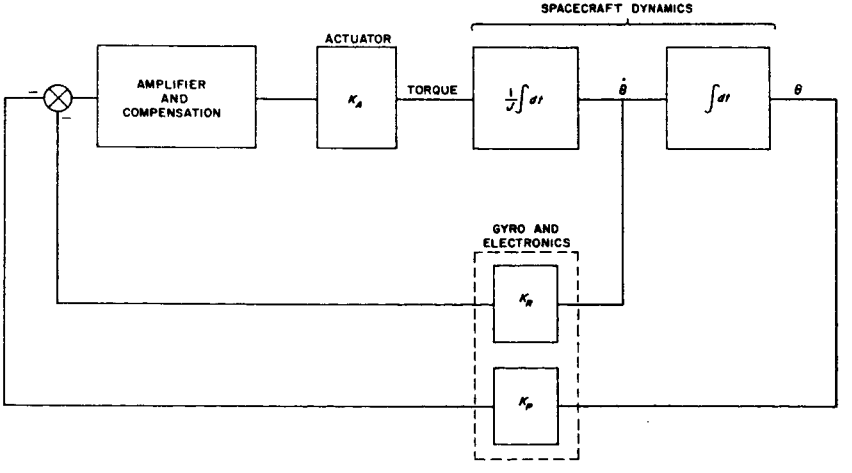


FIGURE 9.2—Simplified autopilot control loop.

ARTICULATION CONTROL SYSTEMS

The final type of control system to be discussed in chapter 9 is that for an articulating member. High-gain antennas, guidance sensors, and scientific-instrument platforms are all examples of devices that often need to be moved with respect to the spacecraft. In many cases this is accomplished by utilizing a closed-loop control system.

These control systems have three essential elements: a sensor, an actuator, and interconnecting electronics. The sensor might be an Earth sensor, planet sensor, or horizon scanner. Typical actuators are standard servomotors or stepper motors. The stepper motor has the advantage that, very often, no velocity feedback is required for damping in the control loop. The disadvantage, however, is that the output motion is not smooth and continuous, but incremental. Since both servomotors and steppers typically have low output torque and high speed, some sort of mechanical speed reducer is used to obtain a torque multiplication and speed reduction.

One possible scheme for pointing scientific instruments at a planet during a flyby mission is shown in figure 9.3. The sensor here is a horizon scanner which detects the local vertical of the planet. The motor-driving electronics act on the error signal to drive a standard servomotor-tachometer combination. The tachometer signal stabilizes the loop. The motor, through the speed reducer, drives the platform to which the instruments are attached.

As in previous discussions, the design of an articulation control system is highly dependent upon the mission and upon other subsystem requirements. The above example should not be considered typical but merely an indication of one mechanization the designer might consider in selecting the final system. Other schemes could be considered and tradeoff studies conducted in a complete evaluation.

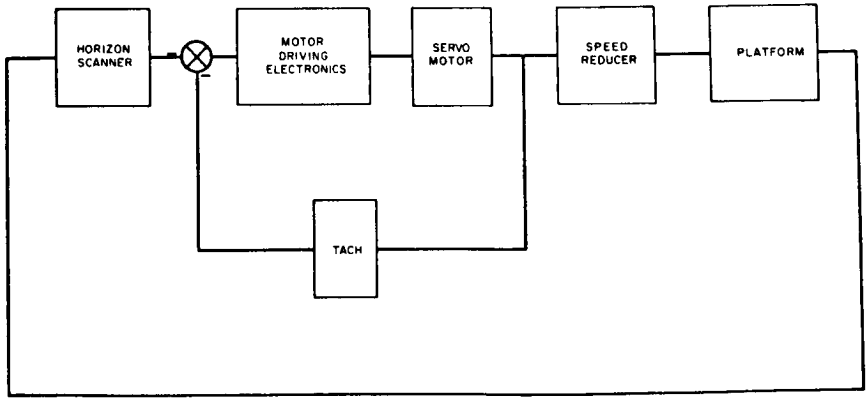


FIGURE 9.3—Articulation control loop.

CHAPTER 10

Inertial Guidance

INERTIAL GUIDANCE (refs. 5–7) is the process of directing the motion of a vehicle from one position to another by sensing the acceleration forces acting on the vehicle in a known coordinate system using instruments which mechanize Newton's laws of motion. These accelerations are then integrated to provide measurements of velocity and distance which are used to guide the vehicle to its destination.

The advantages of inertial guidance and navigation systems are:

(1) Freedom from external influences
(2) Lack of major ground facilities
(3) Invulnerability to countermeasures
(4) Lack of emitted radiation, thus preventing enemy detection
(5) Self-contained, hence independent of communication-link failure.

Each of these advantages provides a different type of benefit, depending on whether the system has to meet military, NASA, manned, or unmanned requirements.

The coordinate system of the acceleration-measuring devices is dependent on both the purpose and the mechanization of the guidance system. Terrestrial applications of inertial guidance, such as submarine and aircraft navigation, commonly use a coordinate system based on local vertical, East and true North. This not only allows direct measurement in terms of latitude and longitude but also allows the use of certain automatic compensation techniques which will be discussed later. Ballistic-missile inertial systems might be optimized by the choice of coordinates in the target direction, cross-range horizontal, and either local vertical or at some angle along the flight path. An inertial guidance system for lunar or interplanetary injection might use a coordinate system oriented along the desired injection velocity vector or any combination of the previously mentioned choices. Some accelerations are measured relative to inertial space, while others are measured in vehicle or airframe coordinates and resolved into the desired reference frame.

Some of the reasons for the choice of various instrument orientations and coordinate systems will be discussed in more detail in the following sections.

STABLE PLATFORMS

The gyros and accelerometers of an inertial-guidance system are normally grouped together in a device known as a stable table, stable platform, or, more simply, a platform. The device derives its name from the fact that the accelerometers are mounted on a base which is gyro stabilized to provide isolation against angular motion of the vehicle. This arrangement permits the accelerometers to measure acceleration forces in a coordinate system which is stabilized in inertial space. It also provides an ideal environment for the gyroscopes, since the precision types used for inertial guidance normally have limited angular freedom and operate with minimum errors when sensing very small angles from a null orientation.

The gyros and accelerometers are mounted in a cluster which is supported with three degrees of freedom relative to the vehicle. This is usually accomplished by means of a set of three or more gimbal rings having orthogonal axes. An attitude error of the stable element in a given direction is sensed by one of the three gyros, and the error signal is amplified and used to drive a gimbal torque motor which returns the stabilized element to null. Each of the three gyros provides stabilization about its own respective axis. The particular gimbal which is torqued by a specific gyro may change as the orientation of the vehicle varies, due to the rotation of the gimbal angles relative to the stabilized element. To compensate for this, the gyro errors are resolved to convert them to gimbal coordinates by sensors mounted on the gimbal pivots. The gimbal axis closest to the stabilized element usually is oriented in the direction requiring the greatest angular freedom. As an example, in a ship navigation system, the inner gimbal axis is usually azimuth. A ballistic missile, on the other hand, might have a pitch inner gimbal if the entire vehicle is oriented in the direction of the target. If unlimited angular freedom is needed about all three axes, a fourth gimbal is required to prevent "gimbal lock," a condition where two gimbal axes become coincident. This configuration is used for fighter aircraft and many space launching vehicles which may be required to maneuver in an unlimited fashion.

Two basic gimbal configurations are in common usage. The conventional arrangement shown in figure 10.1 is one in which the inertial instruments are mounted on an inner cluster surrounded by a group of external concentric gimbals having orthogonal axes. The gimbals can take the form of rings or sections of a sphere or ellipsoid. This configuration allows large or unlimited freedom of all axes. The second arrangement shown in figure 10.2 is one in which the stable element is split into two halves like a dumbbell. The shaft connecting

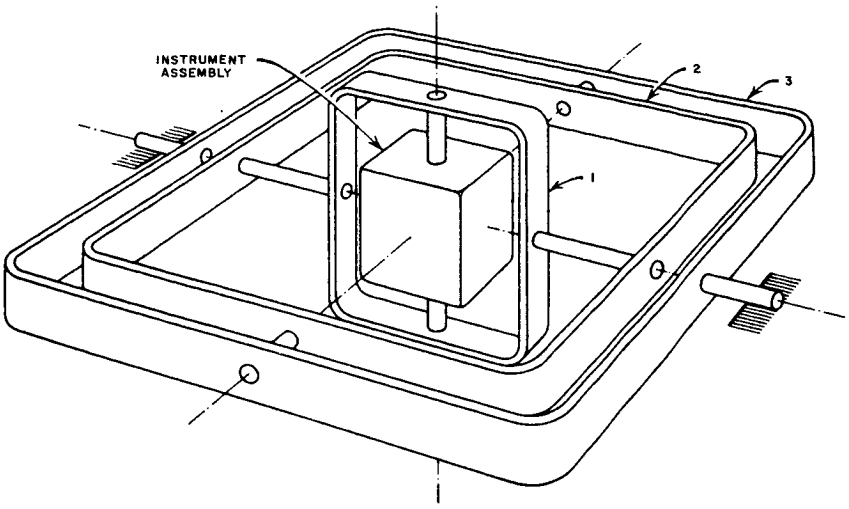


FIGURE 10.1—External gimbal system.

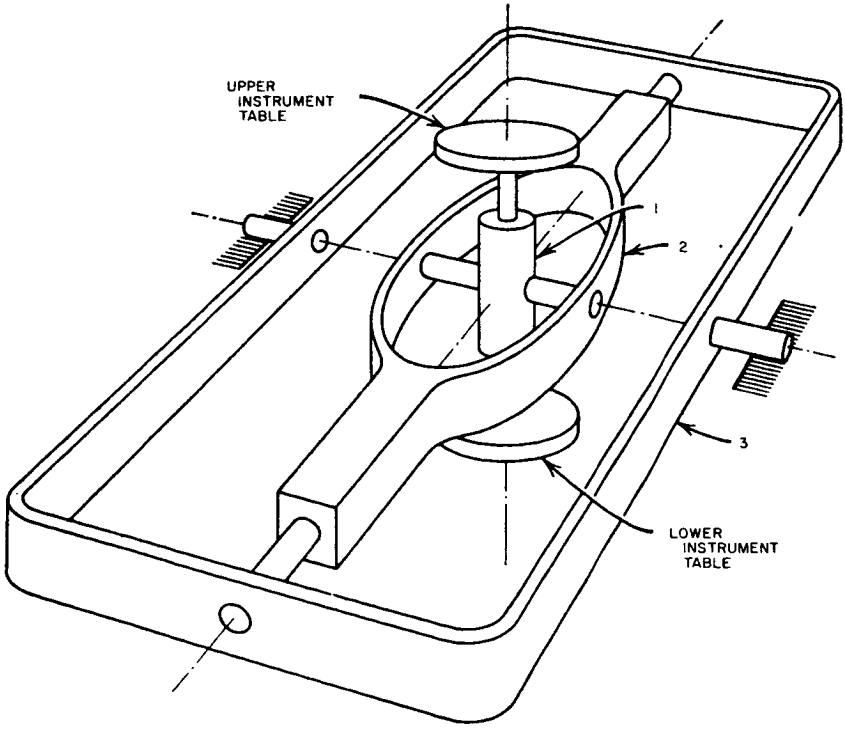


FIGURE 10.2—Internal gimbal system.

the two halves of the assembly passes through a collar and forms the inner axis of the platform. The collar is surrounded by two small yokes or gimbals to allow the other two degrees of freedom. The fact that the instrument assembly protrudes beyond the gimbals restricts the angular freedom of the outer two axes, but allows unlimited rotation of the inner axis shaft in its collar bearing. This arrangement is used for some applications where angular motion of any large amount is restricted to only one axis. It has the advantage of providing very easy accessibility to the instrument cluster for replacement or adjustment. The choice of gimbal configuration is usually based on compactness, form factor, accessibility, expected environment, and the angular freedom requirements.

Additional elements required as a portion of a platform system include gimbal angle pickoffs for use in the guidance and steering computations, sliprings or flexible leads to bring power and signal wires across the gimbal pivots, gimbal torque motors, and leveling or initial orientation sensors. Normally the power supplies, electronic amplifiers, and computer are located as separate, external packages to the platform itself.

STRAPDOWN SYSTEMS

In an alternate arrangement for the configuration of the inertial components of a guidance system, known as a strapdown system, the inertial components are body mounted directly to the vehicle. The accelerometers now measure accelerations in body-fixed or vehicle coordinates. These accelerations must be resolved into an inertial reference frame by coordinate conversion using gyro attitude data. The attitude reference might consist of a three-degree-of-freedom platform containing resolvers on each axis, two two-degree-of-freedom gyros with resolvers, or three single-degree-of-freedom gyros. The coordinate conversion might be done directly using resolvers on the gimbal axes or by processing rate or position data in a digital computer to give an orientation matrix to perform an arithmetic rotation of the acceleration components. Strapdown systems are normally simpler from the instrument standpoint and thus can be made lighter and smaller. They usually suffer from a lack of accuracy or increased computer complexity as compared with a stable platform system.

LEVELING AND ORIENTATION SYSTEMS

To aline an inertial guidance system initially, some arrangement must be provided for orientation of the gyros and accelerometers in the appropriate reference directions. Some of these same alinement

techniques can be used to provide compensation for gyro-drift errors over an extended operating time. Most of these alinement or compensation systems depend on sensing and orienting, relative to the direction of some external reference, such as local gravity, the Earth's polar axis, celestial objects, optical or radio lines of sight, or other platform systems.

The most commonly used method for orientation of two of the three axes of a platform is sensing and alining to the local gravity vertical. Some form of pendulum to sense the vertical has been used for many years to erect or maintain the verticality of gyro axes. The artificial-horizon gyro used as an aircraft instrument contains a pendulous element which keeps the gyro vertical. However, these very simple systems suffer from a serious problem. When the aircraft goes into a turn, the gyro pendulum experiences both gravity and centrifugal accelerations which cause it to aline the spin axis to a spurious vertical. Several methods to counteract this difficulty have been developed. These techniques are largely based on some method for sensing a turning rate above a preset threshold and disabling the pendulum. Although these systems are satisfactory for low-speed aircraft, they cannot be used for high-speed fighter aircraft or missiles.

An idea to solve this problem was advanced in 1923 by Maxmilian Schuler. Schuler suggested that a pendulum having a length equal to the Earth's radius, giving it an 84-minute period, would make it independent of vehicle movement. Schuler pointed out that—

> If the length of the pendulum were equal to the radius of the Earth . . . it would be possible to move the point of suspension around at will on the Earth's surface without disturbing the pendulum [bob] in the slightest. This is because the pendulum's center of gravity would always remain at the center of the Earth, and hence at rest.

Schuler said that although one could not construct such a pendulum, the effect would be obtained with a pendulum system having the same 84-minute period of a string pendulum with a length of the Earth's radius.

The Schuler-tuned pendulum, as it is commonly called, is normally mechanized in a platform system by integrating the output of an accelerometer having its sensitive axis horizontal and using the resulting signal to torque the gyro on that axis. Another approach is that the integrator output being proportional to vehicle velocity causes the platform (and thus the accelerometers) to rotate continuously at a rate which is the same as the angular velocity of the vehicle around the Earth. The action of a Schuler-tuned system is like an undamped pendulum. Although the inertial system will seek out and maintain the local vertical, and instrument errors will not cause

238-635 O—67——8

an angular error to build up with time, the 84-minute pendulum will oscillate at an amplitude bounded by errors in the instruments. Damping is quite often provided by introducing an external velocity correction from Doppler radar or some other source in order to reduce the oscillatory offsets. If the leveling system is only to be used for initial orientation, the loop gains are normally increased to shorten the erection time, and damping is then usually provided by the introduction of compensation networks in the amplifiers.

The third platform axis, azimuth, is normally alined by one of two methods. For land or aircraft navigation systems used over long-time durations, gyrocompassing is employed. The fact that the Earth rotates in inertial space about its polar axis can be sensed by an inertial platform. By orienting the azimuth gyro so that its spin axis is in the North-vertical plane and the input axis is East or West, any deviation of the input axis from the true East-West direction will introduce a component of Earth's rate. This rate can be sensed and used to torque the platform so that the gyro remains on an East-West heading. The accuracy of the gyrocompassing is determined by gyro errors and the latitude of the vehicle. At high North or South latitudes, the horizontal component of Earth's rate becomes very small and the gyro errors produce a larger deviation from true North. Gyrocompassing, like Schuler tuning, may be used both for initial alinement as well as compensation or trimming during moderate vehicle motion. It suffers a similar drawback, however, in that it cannot be used under conditions during which the vehicle is exposed to a high acceleration or rotation environment.

For initial alinement of platforms requiring very precise azimuth accuracy, an optical orientation technique is used. Collimated light from a precisely oriented source is beamed to a Porro prism mounted on the azimuth gimbal of the platform. The reflected beam from the Porro prism returns to the orientation station where it passes into a detector. Any angular deviation in azimuth of the platform causes a change in the angle of the return beam. The output of the detector can thus be used to command the azimuth axis of the platform to a position where the plane of the Porro prism is perpendicular to the source beam, causing the return beam to be collinear. Angular errors as small as 1 second of arc can be measured in this manner.

INERTIAL-GUIDANCE APPLICATIONS

Inertial systems may be used for two primary applications, navigation and guidance. The navigation task is essentially to provide a continuous position fix relative to some predetermined reference coordinate system. The guidance task is to provide steering and

velocity control signals to the vehicle for guiding it to its preselected destination.

The navigation task is a relatively simple one for an inertial-guidance system. The basic function is accomplished by double integration of the acceleration components to obtain the position change from the starting point. The task is made more complex if the vehicle traverses a large portion of the Earth's surface, because the Earth is not a perfect sphere and is rotating in inertial space. Errors between the plumbline vertical and the perpendicular to the horizon can reach 11 minutes of arc at 45° latitude. The Coriolis effect illustrated in figure 10.3 introduces a "phantom" acceleration into the system, depending on the angle of the vehicle's path with respect to the polar axis. However, the navigation task is fundamentally one of bookkeeping.

The guidance task is somewhat more complex, since not only current position but also velocity must be known and used to obtain the guidance signals. In addition, errors introduced by thrust and center-of-gravity misalinements, as well as wind or tide disturbances, must be compensated for.

Guidance systems may be divided into two classes, depending upon the form taken by the guidance equations: explicit, or implicit. In the explicit class, the equations of motion of the vehicle are mechanized in the computer, and the computer solves these equations during vehicle motion. The solution of the equations is used continuously to determine the deviation from the desired directional path and velocity profile and to generate commands to null this error, either instantly or at the expected time of arrival.

The implicit or delta type of guidance compares the actual vehicle motion with a predetermined or standard trajectory. The guidance equations are formulated in terms of differences or deltas between various measured quantities and their standard values. In many implicit guidance systems, the difference quantities are expanded in a power series to determine the steering or velocity control signals.

The exact details of mechanization of an inertial guidance system are very dependent on the mission requirements. As a result, it is not possible to describe a configuration which would satisfy all applications. However, in almost all configurations the limiting factor determining the accuracy of the system is not the guidance equation mechanization, but the basic inertial components themselves.

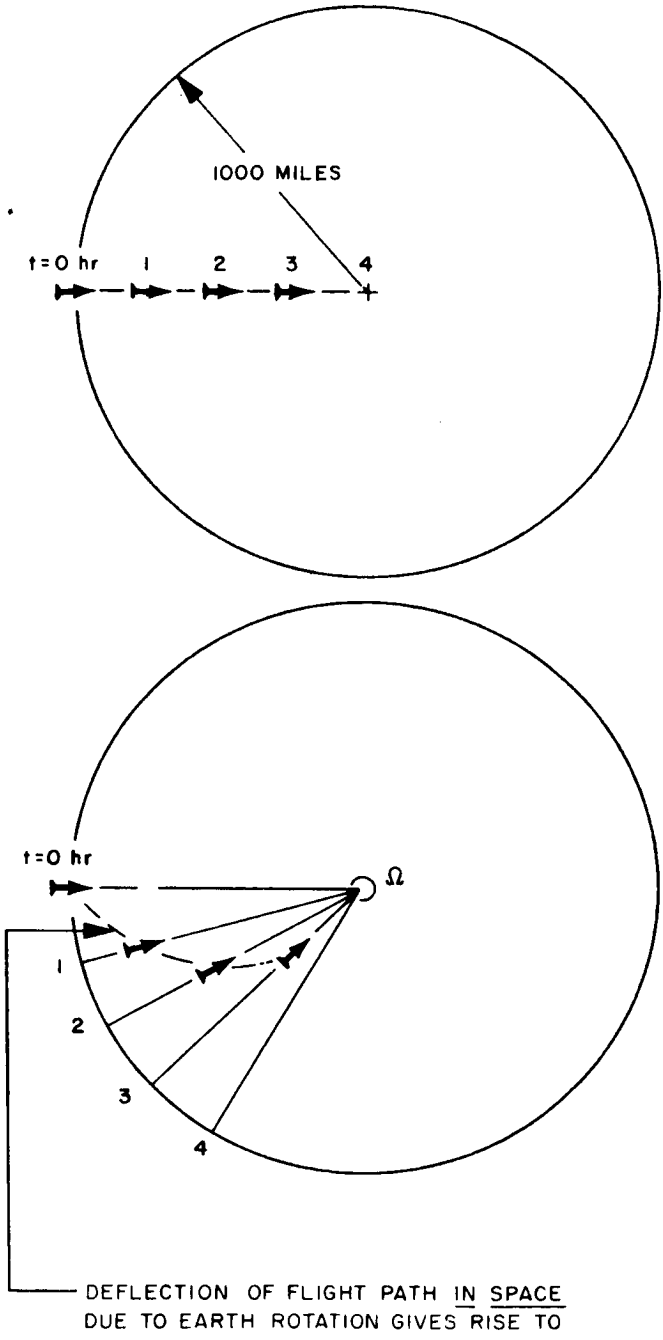


FIGURE 10.3—The Coriolis effect.

CHAPTER 11

Earth-Based Midcourse Guidance

IN THIS SECTION, the guidance of unmanned lunar and planetary spacecraft by means of Earth-based radio tracking and command will be discussed (see ref. 12).

A practical axiom in space guidance is that any function which can be performed equally well on the Earth, or in the spacecraft, should be performed on the Earth. Since unmanned spacecraft must be tracked from the Earth to receive information from them, tracking data are readily available. Also, most spacecraft will possess an attitude sensing and control system, by means of which the spacecraft can be pointed in any desired direction. Thus, with the addition of a command link to the spacecraft and a propulsion system, the latter needed for guidance in any case, a guidance system is created. The capabilities of this guidance system will be examined in the following discussion.

TECHNICAL DISCUSSION

In figure 11.1, the assumed guidance system is shown schematically. Three tracking stations are assumed, to provide continuous coverage. Tracking data are fed to a centrally located computer (fig. 11.1a). Typical tracking data would be azimuth and elevation angles, range rate, and possibly range. The orbit is determined, and a command is transmitted to the spacecraft (fig. 11.1b); the command consists of the direction and magnitude of the corrective velocity maneuver required to adjust the trajectory. In figure 11.1c, the spacecraft executes the maneuver.

The accuracy at the target, which may be Mars or possibly a given crater on the Moon, will be affected by two almost independent error sources: (1) the error in the determination of the orbit based on the tracking data, and (2) the error committed in the execution of the maneuver.

The accuracy of the orbit determination is affected by errors in the tracking data, errors in our knowledge of the locations of the tracking stations, errors in certain physical constants such as the velocity of propagation and the Astronomical Unit (AU), and errors in computation. The mathematical procedure for orbit determination which

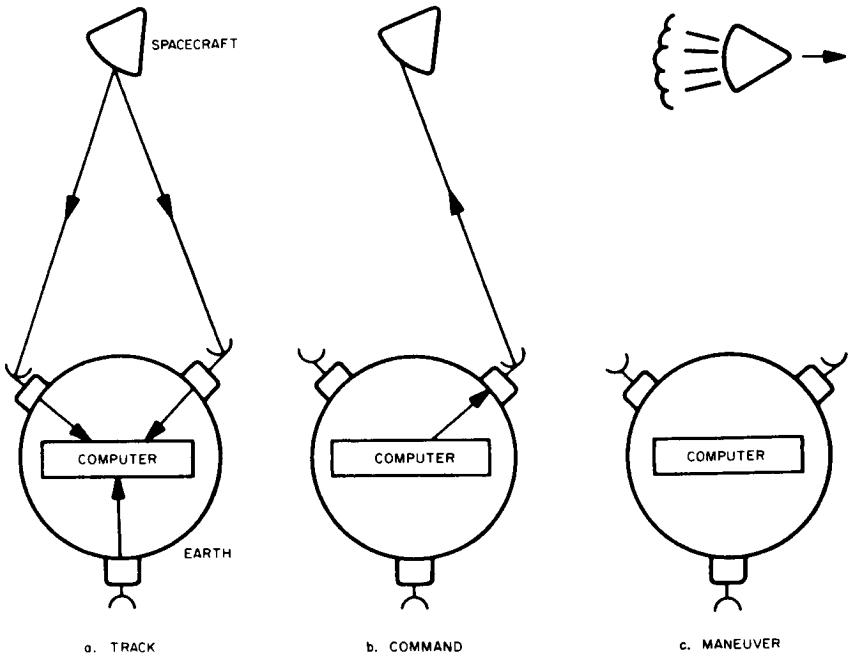


FIGURE 11.1—Guidance-system schematic.

has been found most effective for lunar and interplanetary flight is based on the maximum-likelihood method of statistical estimation, which is, in effect, a least-squares fit of the entire orbit; in this procedure, the fact that the spacecraft must obey Newton's laws, which are precisely known, is fully utilized.

Figure 11.2 shows an example of the accuracy of orbit determination for a lunar trajectory. Angular measurements accurate to approximately 2 mrad and range-rate measurements accurate to 0.15 m/sec were assumed. The miss component, which is the semimajor axis of the 40-percent probability ellipse, may be interpreted as a sort of circular probable error. The horizontal portion of the curve corresponds to a time at which it was assumed that the probe was not visible from a tracking station and, hence, no new data were being received. Figure 11.2 shows that the first few hours of tracking data are especially powerful in determining the orbit. Also, as expected, the curve decreases monotonically downward, since additional data must always improve the accuracy, however slightly. Finally, we see that we can determine, with an accuracy of better than 10 kilometers, where the spacecraft will impact on the Moon. Corresponding accuracy for Mars and Venus would be approximately 1000 kilometers.

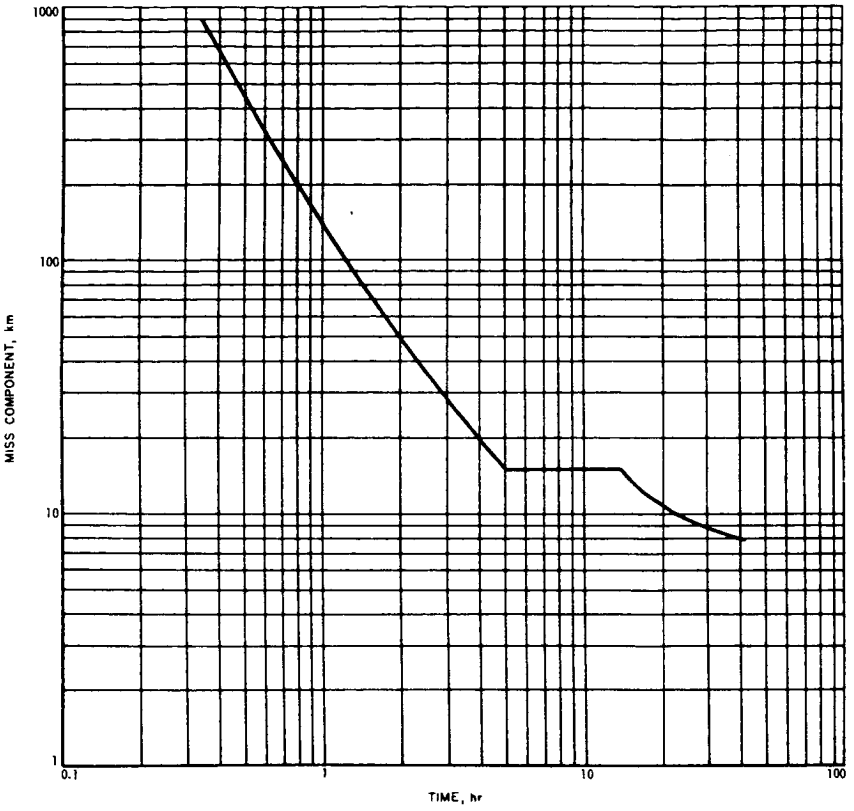


FIGURE 11.2—Lunar orbit determination accuracy.

If we could execute a midcourse maneuver perfectly, then the accuracy of figure 11.2 could be achieved. However, since errors will be made in the execution of a midcourse maneuver, and since we wish to minimize the magnitude of the maneuver while at the same time minimizing the miss distance at the target, some analysis of the maneuver will be necessary.

In figure 11.3 (a), a midcourse maneuver is represented. The maneuver ΔV is assumed to be executed instantaneously, and the orientation of the $V_1V_2V_3$ coordinate system is arbitrary. In figure 11.3(b), the target geometry is shown. The $M_1M_2M_3$ coordinate system is centered at the target and moves with the target. The M_3 axis is taken as the direction of the standard or error-free trajectory. The actual trajectory is assumed to be parallel to the standard trajectory (a good assumption). The miss components, m_1 and m_2, will be available from the orbit determination. The small shaded area in figure 11.3(b)

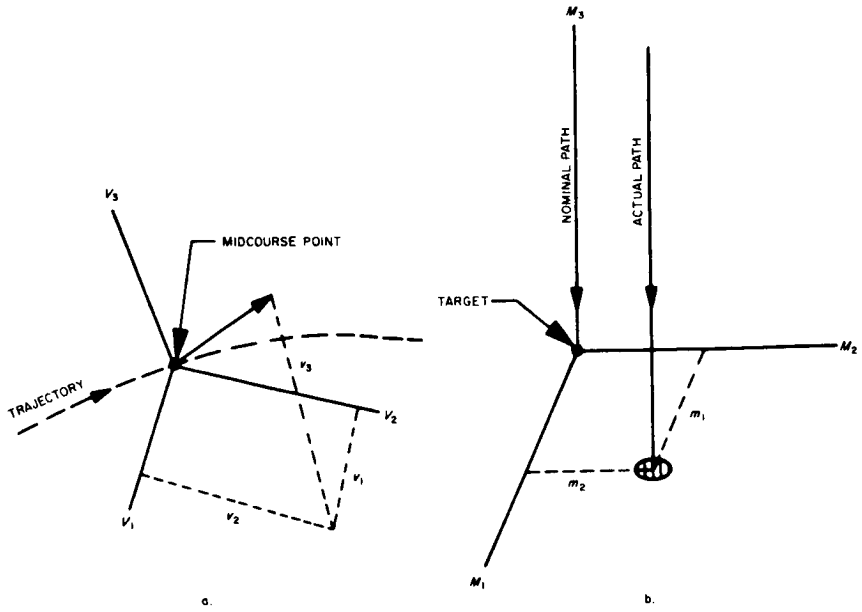


FIGURE 11.3—Midcourse maneuver and target geometry.

indicates the uncertainty in m_1 and m_2. Let us next define the sensitivity coefficients:

$$\lambda_{ij}=\frac{\partial m_i}{\partial v_j} \tag{11.1}$$

Since we wish to hit the target, we can then write

$$\left.\begin{aligned} m_1'=-m_1=\sum_{j=1}^{3}\lambda_{1j}v_j \\ m_2'=-m_2=\sum_{j=1}^{3}\lambda_{2j}v_j \end{aligned}\right\} \tag{11.2}$$

Observe, however, that we have two equations and three unknowns, the three unknowns being v_1, v_2, and v_3. Thus, one additional degree of freedom remains, resulting from the fact that the $M_1M_2M_3$ coordinate system moves with the target, and we have left free the time at which the spacecraft hits the target. The remaining degree of freedom could be used to control the time of flight or to minimize the magnitude of the midcourse maneuver.

It now remains to determine the point at which the midcourse maneuver should be applied. The magnitude of the midcourse maneuver will be approximately proportional to the error at the injection point.

However, for a given injection error, there are three conflicting criteria between which a compromise must be made:

(1) The accuracy of orbit determination (see fig. 11.2) improves as time increases, suggesting that the maneuver should be made as late as possible.

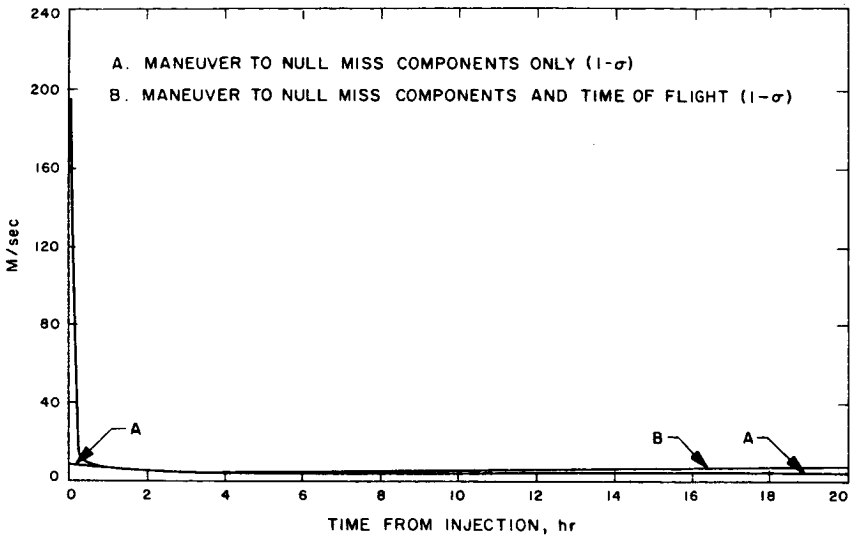


FIGURE 11.4—Magnitude of maneuver versus time.

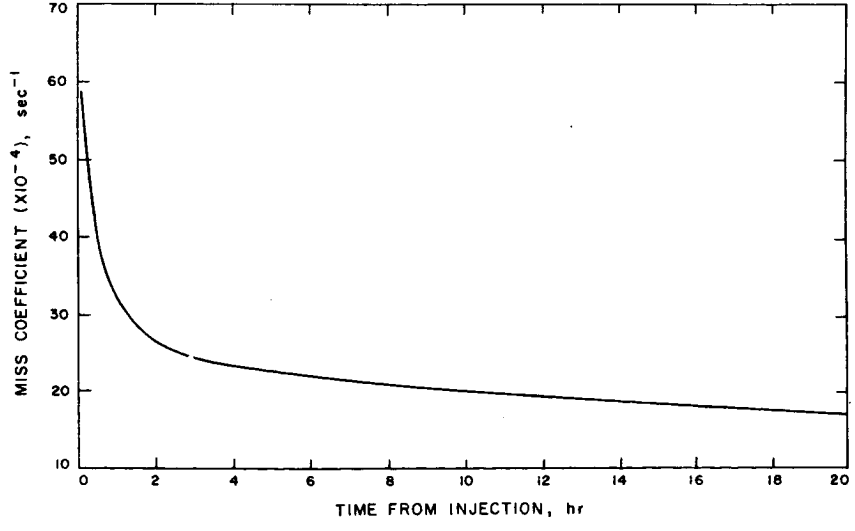


FIGURE 11.5—Miss coefficient versus time.

(2) The magnitude of the required maneuver increases as time increases (see fig. 11.4), suggesting that the maneuver should be made as early as possible.
(3) The sensitivity of errors at the target to errors made during the midcourse maneuver decreases with time (see fig. 11.5).

Fortunately, we find that the tradeoff between the above criteria yields broad optima, and the time of making midcourse maneuver is not critical; for lunar trajectories, a time of 10 to 20 hours after injection is acceptable, and for interplanetary trajectories a time of from 1 to 3 days after injection is satisfactory.

Finally, in table 11.1 certain key features of midcourse guidance systems are summarized. The figures given in table 11.1 should be interpreted as being suggestive of the general region in which the quantities lie, rather than precise data.

TABLE 11.1—Midcourse Guidance Performance[a]

Target	(1) Miss due to representative injection guidance, km	(2) Assumed tracking accuracy		(3) Orbit-determination accuracy from (2), km	(4) Midcourse maneuver to correct (1), m/sec	(5) Accuracy of maneuver (assumed)		(6) Error due to maneuver, km	(7) Total accuracy rms of (3) and (6), km
		Radius	m/sec			Pointing angle, deg	Magnitude, percent		
Moon	6 000	2×10^{-3}	0.15	10	40	½	1	64	65
Mars	500 000	2×10^{-3}	.15	2500	20	½	1	5400	6000
Venus	300 000	2×10^{-3}	.15	1000	20	½	1	2700	2900

[a] All quantities are 1σ.

CHAPTER 12

Celestial Navigation

THIS SECTION WILL EXAMINE the requirements for an interplanetary celestial navigator and describe several approaches for mechanizing the basic optical sightings required to determine the orientation and position of the space vehicle (ref. 12).

In the previous section it was seen that Earth-based radio midcourse guidance would be adequate to deliver a spacecraft to the vicinity of the near planets. If the mission requires closer proximity than is provided by radio midcourse guidance, an approach maneuver made within 1 to 2 million kilometers of the destination planet could reduce the miss distance to a satisfactory level. What are the reasons for the development of a self-contained celestial navigator?

Advanced interplanetary missions will impose additional guidance requirements on the spacecraft. Celestial navigation will undoubtedly be needed for such missions as return trips from the planets, trips to the more distant planets, military missions where communication security is required, and to provide redundancy for manned missions. A celestial navigator may prove to be optimum for extremely accurate planetary flights where the same instruments might serve the dual purpose of midcourse and terminal sightings.

TECHNICAL DISCUSSION

Many techniques for celestial space navigation have been proposed. Various geometrical relationships have been suggested, including location by intersections of conical surfaces, planes, lines, and assorted figures of revolution. Basically, most systems boil down to a position fix determined by the angles measured between two stars and two bodies of our solar system. An example of the related geometry is shown in figure 12.1. The angle between star 1 and planet A locates the spacecraft on a circular cone with cone half angle α_1. The angle between star 2 and planet A similarly locates the spacecraft on the cone having axis AS_2. The two cones intersect A in two straight lines. The ambiguity can be removed by the gross knowledge of the spacecraft's location in the solar system. Similar sightings on planet B will locate the spacecraft on a line

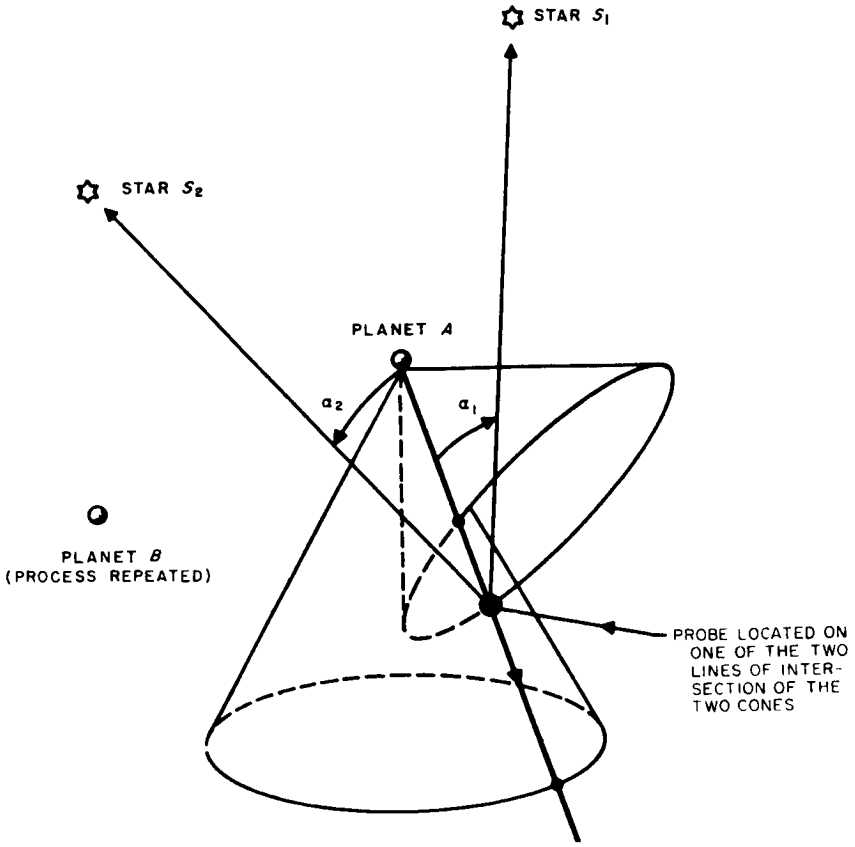


FIGURE 12.1—Related geometry for celestial space navigation.

through B which intersects the line from A at the location of the vehicle.

To determine the trajectory which the spacecraft is following, the velocity also must be determined. Direct velocity measurement by stellar spectral shift or relativistic methods has been suggested by several investigators, but this method does not appear to hold promise of accuracies of better than 100 m/sec because of fluctuations or nonuniformity of the starlight. The most suitable method proposed is the taking of two or more position fixes at a known time interval. For midcourse guidance, the point of injection could be used as one of the position fixes.

Three general classes of optical measurements can be applied to celestial navigation. Absolute measurements of the angles between the planets and stars could be made by several independent star trackers gimbaled relative to each other. Relative measurements of

the planets with respect to the star background could be made by correlation or map-matching techniques. Measurements of one star or planet at a time could be made by a single optical sensor relative to an inertial platform or by time measurements in a vehicle rotating at a uniform speed.

Each of the three classes of optical measurements has advantages and disadvantages. The system employing several independent trackers can make all angle measurements at the same time, thus simplifying the job of the computer. This system suffers from the standpoint of weight and complexity of the trackers, and the accuracy is dependent on precise spacecraft attitude control. The requirements on the attitude-control system could be reduced by adding the instantaneous error of the trackers to the indicated gimbal readout position at the expense of complicating the data processing.

The second method of measurement relative to the star background is less dependent on precision attitude control and uses fewer optical trackers. This system has a disadvantage in that it must track the planet relative to stars of perhaps 8th or 10th magnitude. A tradeoff must be made between having a large field of view to track brighter stars but at reduced accuracy and improving the accuracy by reducing the field of view and requiring the use of relatively dim stars.

Systems in the third class, using a platform-stabilized tracker or a spinning wide-angle camera with a number of slits in the focal plane of the detector, have an advantage in that they utilize only a single tracking device. In addition, the platform-stabilized tracker has the advantage of being independent of spacecraft motion, but the disadvantage is that a complex stable-platform tracker has the same problem as the sensors in the first class, in that it must measure large angles to great precision. The spin-scan tracker is simple and would not require moving parts if the entire vehicle were spin stabilized. In addition, time rather than angle measurements would be used, thus simplifying the data processing. The spin-scan system has a serious drawback in that it is subject to errors caused by nutation and precession motions of the spinning vehicle. Motions of this type rendered almost impossible the interpretation of the data from the spin-scan cloud-cover satellite, Vanguard II. Further difficulties result from a spinning spacecraft, since orientation of solar panels toward the Sun and a directional communications antenna toward the Earth cannot be achieved at the same time. Examples of several celestial navigator schemes suggested by the previous discussion are shown in figure 12.2.

To achieve the same position and rate accuracies shown for radio midcourse guidance in table 11.1, the celestial navigator would be

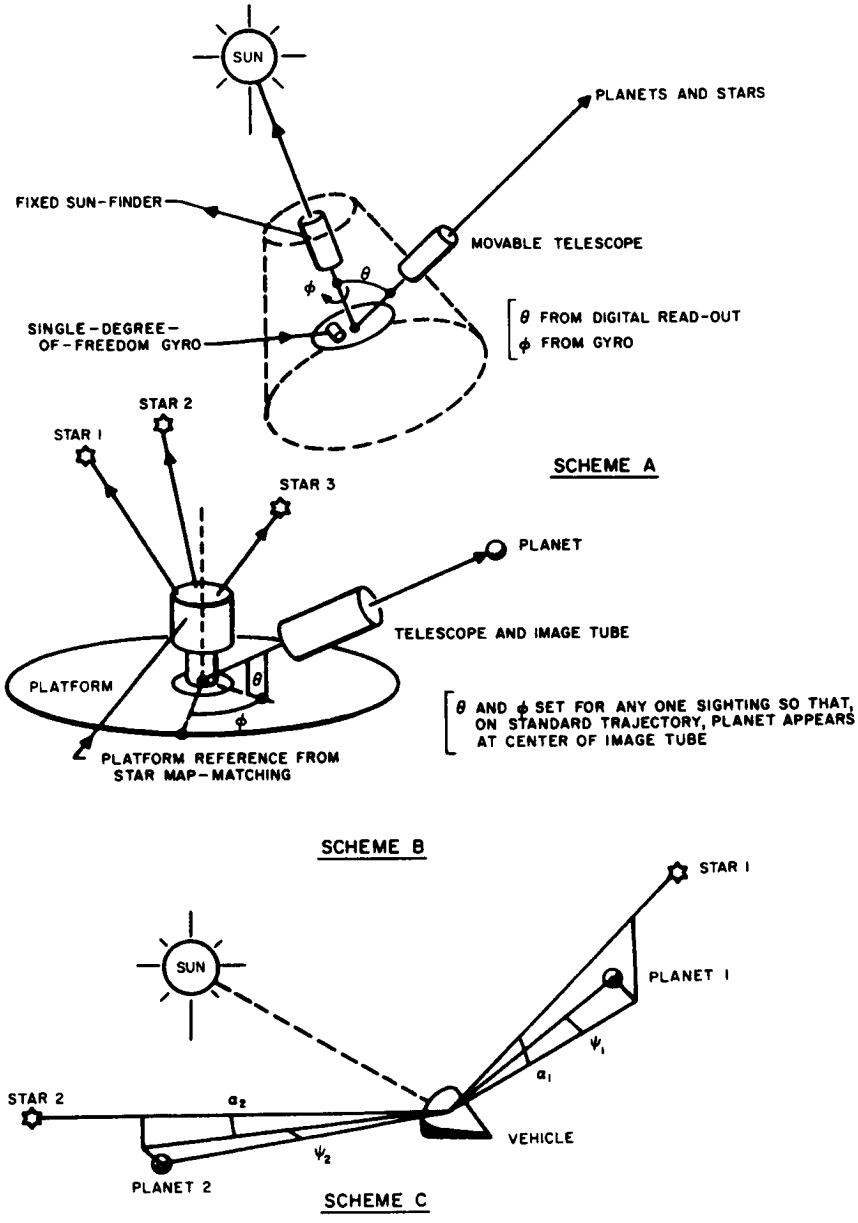


FIGURE 12.2—General celestial navigator schemes.

required to measure angles to approximately 1 second of arc at two points widely spaced on the trajectory. The celestial navigation accuracy is dependent on the amount of a priori data used

and the processing techniques applied. An optimum system would make use of Kepler's laws and would perform least-squares smoothing of the measurements. Instead of actually computing the spacecraft's trajectory on the basis of an ephemeris of the stars and planets, the computer would probably measure the difference between angles which should occur on a standard trajectory and those actually observed. Using linear perturbation theory, the differences could be operated on by a stored-program digital computer to determine the required maneuver.

The point at which the maneuver should be made is subject to the same constraints as Earth-based midcourse guidance, discussed in the previous section. The accuracy of the trajectory relative to the destination planet improves with time after injection, and the required sighting accuracy is reduced. At the same time, the sensitivity to midcourse maneuver errors decreases with time. The magnitude of the maneuver required to remove the injection guidance errors increases with time. Since the celestial navigator requires a longer time to obtain sufficiently accurate information on which to base a midcourse maneuver than an Earth-based system, it will require a larger velocity increment and thus have a heavier correcting rocket. On a planetary trajectory, a velocity increment of more than 100 m/sec would be required to correct for a 500 000-kilometer miss of Mars if the maneuver were made past the midpoint of the trajectory.

CONCLUSIONS

Celestial navigators may be mechanized in a variety of ways depending on the configuration of the spacecraft and the mission. A property in common to each of the methods is that the angles between at least two stars and two bodies of the solar system must be measured to great precision. Spacecraft velocity can be measured only by two such position fixes made as far apart as maneuver fuel economy permits. Most missions to the Moon or near planets can be easily handled by Earth-based midcourse guidance, but a celestial navigator might be required for advanced interplanetary missions.

238-635 O—67——9

CHAPTER 13

Lunar-Landing Guidance

THIS SECTION WILL DEAL with requirements and techniques for the descent and soft landing of an unmanned spacecraft on the Moon (see ref. 12). The magnitude of the arrival velocity of the spacecraft in the vicinity of the Moon for typical lunar trajectories is of the order of 2600 m/sec, or about 5800 mi/hr. The spacecraft must be oriented and braked from this velocity to arrive at the lunar surface with essentially zero velocity and to land vertically. More practically, however, landing velocity will not be zero but will be limited to a few meters per second in both vertical and lateral directions (fig. 13.1). The spacecraft may incorporate a landing structure to

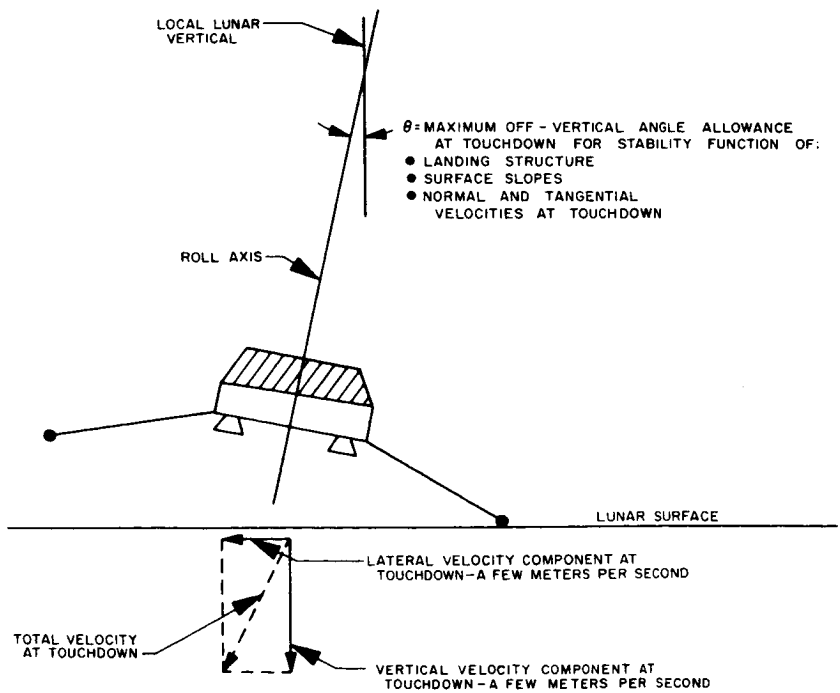


FIGURE 13.1—Landing velocity in both vertical and lateral directions.

minimize shock and the probability of spacecraft toppling during landing. With proper attention to landing dynamics and structure, the touchdown shock can be limited to 25 g or less. By comparison with these conditions, it is interesting to note that a parachutist landing on Earth will normally impact at about 5 m/sec and suffer a shock of 4 or 5 g.

FUNCTIONAL REQUIREMENTS

There are several functional requirements to be met by the flight-control system. First, it must incorporate a propulsion system, appropriately scaled and capable of being modulated in thrust for control of descent velocity. Second, there must be on board the spacecraft an accurate means of determining position and velocity relative to the lunar surface in order to provide the signals for control of attitude and velocity of the spacecraft during descent. Finally, the efficiency of the descent must be kept as high as possible to minimize propellant requirements. Under the best conditions, a lunar soft landing requires a high propellant mass fraction. To reduce arrival velocity to essentially zero, using a propellant with a nominal specific impulse of 300 seconds, about 60 percent of the total weight of the spacecraft, prior to firing of the retrorockets, will have to be allotted to propellant. Only a rigorous optimization of the descent system will insure that no unnecessary penalties are being taken in total propellant utilization.

LUNAR APPROACH

When the spacecraft arrives in the vicinity of the Moon, Earth-based tracking will yield knowledge of its velocity magnitude within a few meters per second and its direction within a fraction of a degree. However, spacecraft position relative to the lunar surface may be uncertain by 10 kilometers or more.

Attitude reference for the spacecraft can be provided, during Earth-Moon transfer, by optical sensors locked to the Sun and Earth (or star) directions. A set of orthogonally mounted gyros may provide rate control about the three spacecraft axes and the reference for short-term precision maneuvering of the spacecraft for thrust-axis alinement. The latter can be accomplished by gyro torquing to the computed direction by Earth-based radio command. Implicit in the spacecraft mechanization is a system of reaction torquing for the continuous maintenance of spacecraft attitude.

When the spacecraft is still 10 to 20 minutes away from lunar impact (corresponding to an altitude of about 1500 kilometers) and still possesses its transfer attitude references and orientation, a ma-

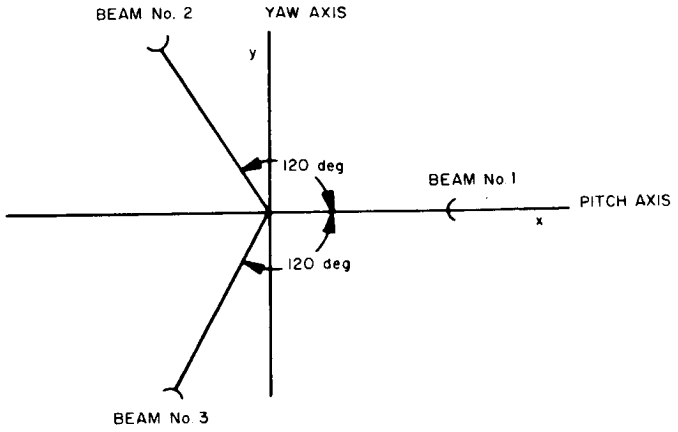
neuver can be initiated by radio command to aline the spacecraft thrust axis opposite to its velocity vector, preparatory to braking the spacecraft for landing. Once descent has been initiated, the spacecraft must continue on its own devices, without further dependence on celestial references or aid from ground tracking or command. Tracking data are not only inadequate for the accuracies required in determination of altitude h and vertical velocity $\dot{h}$, but the time delay in transmission of guidance data between Earth and spacecraft would be prohibitive.

TERMINAL SENSING

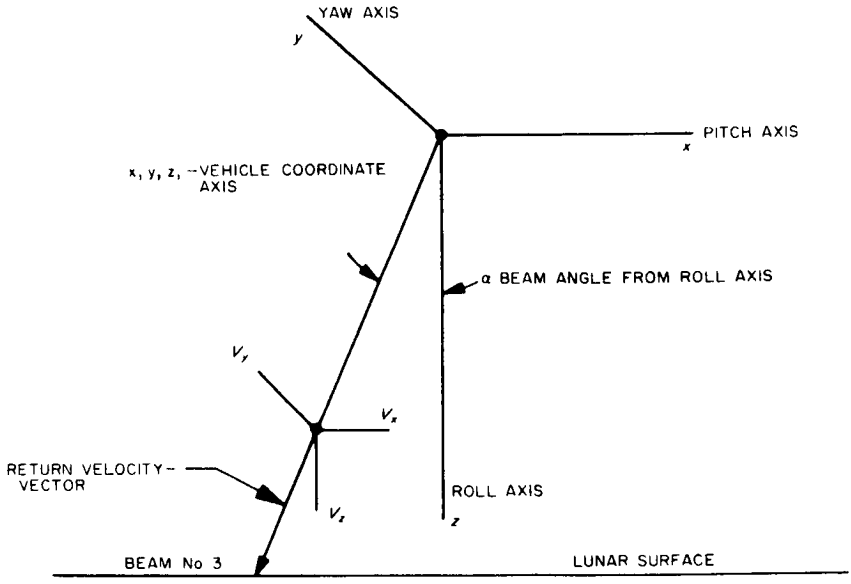
Measurements of position and velocity relative to the lunar surface must be made. For initial ignition of the retro system, a simple measurement of altitude may suffice. Such a measurement may be made either by means of an optical horizon scanner or by using pulse-radar techniques. The choice depends on the altitude of ignition and the required accuracy, and on how the device is otherwise to be employed during descent.

Electrical power requirements and total weight become restrictive for radar equipment at ranges of 100 miles or more. Primary restrictions on the use of optics are the dependence on lunar lighting conditions and the possibility that lunar-surface irregularities could present a "false horizon" to the optics at lower altitudes. Under suitable conditions, either system is capable of indicating ignition altitude to 2 percent or better.

At some point in the descent, measurement of altitude and both vertical and horizontal components of velocity must be initiated, and continued essentially to touchdown. A precision radar altimeter appears to be the best means for the continuous and precise measurement of altitude. Velocity can be determined by means of a multibeam Doppler radar system or by using optical V/H driftmeter techniques. In principle, either a Doppler or an optical system must consist of at least three active beams suitably distributed about the roll axis of the spacecraft and "squinting" out at equal radiating or viewing angles from the roll axis, as shown in figure 13.2. Thus, the signal resolved from each beam (for either system) will be proportional to the relative velocity of the spacecraft with respect to the Moon along the beam direction. By resolution of the velocity information of each beam into components along the pitch, yaw, and roll axes of the spacecraft, error signals for control of both thrust and spacecraft attitude are derived. Again, a disadvantage of the optical system is its dependence on lighting conditions.



a. VELOCITY-SENSOR LOCATIONS-PLANAR VIEW



b. METHOD OF RESOLVING VELOCITY RETURNS

FIGURE 13.2—Geometry of a Doppler or optical system.

RETRO REQUIREMENTS

For maximum efficiency and simplicity, it would appear that the total velocity decrement of 2600 m/sec required for soft landing should be achieved in a single short impulse, at a level of thrust which is limited only by the load capability of the spacecraft, and with a

simple propulsion system, such as a solid rocket. Practically, however, uncertainties in measured position and velocity, in angular alinement of the thrust axis, and in burning characteristics of the engine(s) (as a function of such factors as temperature, mixture ratio, chemical constitution, propellant loading, burning time) would result in altitude and velocity dispersions far in excess of those admissible as soft-landing conditions. It is thus evident that some form of velocity or thrust-level control must be employed in the soft-landing descent. It is also apparent in the mechanization that the descent time cannot be less than the time necessary to generate and execute the necessary guidance commands. Further, engine throttlability requirements should not become excessive.

Within these requirements, two modes of descent present themselves. One utilizes a single engine which is modulated in thrust. The modulation may be partial, say to 90 percent of full thrust, or total, to zero, in which case the engine is started several times. Difficulties in sensing through the flame may require that the engine be turned off occasionally.

The second mode of descent, to be described more fully, employs two separate engine systems, one to provide a fixed increment of thrust at high level for rapid removal of the bulk of the energy, and a lower level, throttlable system for reduction of the remainder of the velocity.

TWO-LEVEL SYSTEM CHARACTERISTICS

The main engine should be sized to remove a major fraction of the spacecraft velocity with a short, fixed-direction impulse at near-constant thrust. Upon termination of this thrust, the vernier system takes over to remove the nominal residual velocity, plus whatever velocity dispersions have accumulated. We assume that the main-engine burning is guided without external sensing, in the event that the exhaust is opaque; we further assume that the vernier system exhaust is transparent, and that full guidance based on external sensing will occur during the vernier period.

For maximum descent efficiency, the main engine must be characterized by a high thrust-to-weight ratio, a high specific impulse, and a high total impulse. On a lower level, the vernier system should be similarly endowed, but, in addition, it must also be capable of being modulated in thrust from T_{max} to T_{min} over a moderate ratio. The minimum thrust of the vernier system should be less than the lunar weight of the spacecraft near propellant depletion to permit the spacecraft to "fall" toward the Moon while still under minimum thrust for attitude control. The maximum thrust of the vernier system

should be as high as practical to maximize the braking rate of the spacecraft during the maximum-effort phase of the terminal descent. The maximum- and minimum-thrust vernier trajectories are shown in figure 13.3.

MAIN RETRO TRAJECTORY

The overall descent profile for the two-level retro system is shown in figure 13.3. This consists of the fixed-impulse, fixed-direction main retro trajectory, with associated dispersions due to initial conditions of descent and engine burning characteristics. The profile also shows the guided vernier descent trajectory to touchdown.

The initial velocity dispersion δV_a at mean-engine ignition is of the order of a few meters per second as derived from Earth-based tracking, which results in a contribution to the altitude uncertainty at main-engine burnout. The initial attitude uncertainty δh_a is directly attributable to the accuracy of an altitude-marking device aboard the

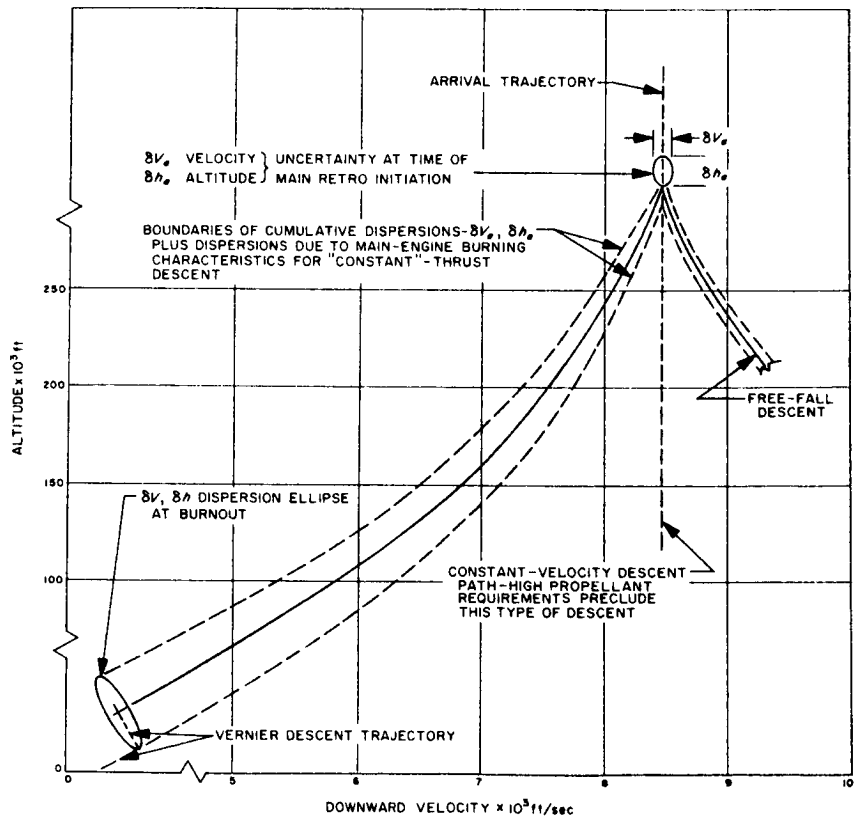


FIGURE 13.3—Overall descent profile for two-level retro system.

spacecraft. A primary effect of δh_a is an increase in velocity uncertainty at burnout.

The third uncertainty which exists at the initiation of descent is in the vehicle thrust-axis alinement, which is fixed inertially for the period of main-engine burn. Some typical alinement geometries are shown in figure 13.4, illustrating that an initial alinement uncertainty of as little as one degree can result in many degrees of uncertainty in the direction of the residual velocity vector at burnout and as much as 100 ft/sec in magnitude. An efficient descent program requires that all such dispersions be minimized.

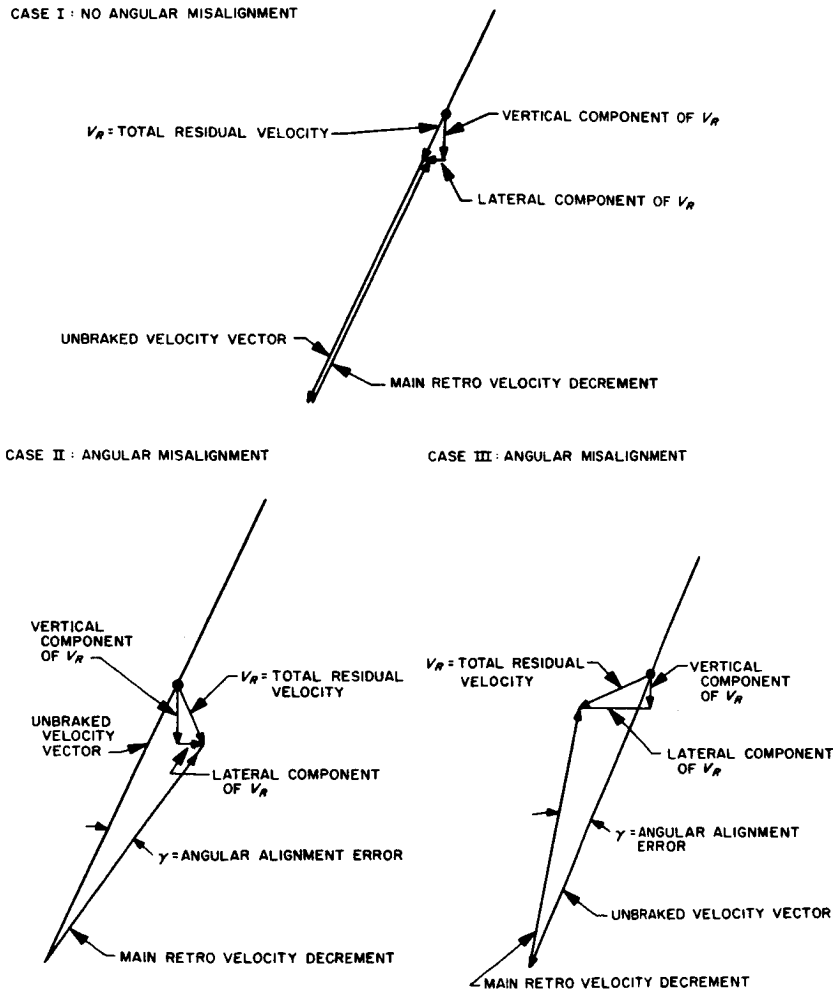


FIGURE 13.4—Some typical alinement geometries.

VERNIER DESCENT

Figure 13.5 is a larger scale profile of the vernier-guided descent trajectory. Vehicle thrust and attitude are controlled during this phase, requiring the previously indicated continuous measurement of altitude and both components of velocity.

At burnout of the main engine (after which it may be jettisoned, if practical, for fuel conservation), the resultant velocity of the spacecraft will, in general, have both lateral and vertical components. However, as a consequence of the velocity dispersions (as shown in fig. 13.5) the velocity vector and the vehicle thrust axis are no longer necessarily alined.

The descent guidance consists of two modes. For maximum efficiency, we shall assume that first there occurs a period of descent from point of main-engine burnout under minimum vernier thrust to a point $(h_t, \dot{h}_t)$ on the h–$\dot{h}$ profile at which maximum effort is required

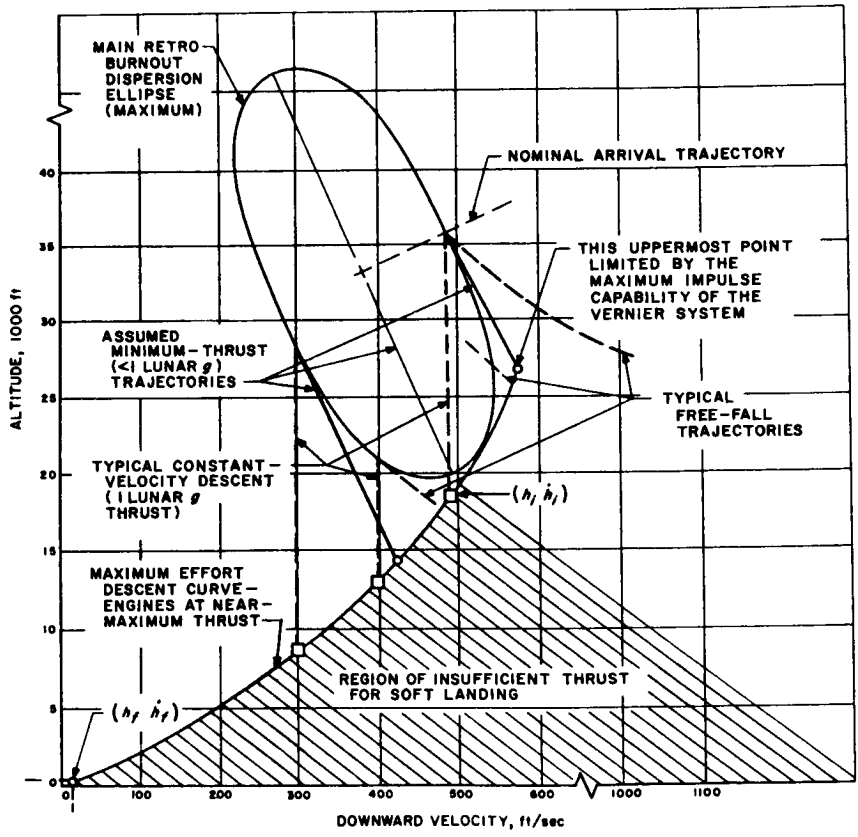


FIGURE 13.5—Vernier-guided descent trajectory.

of the vernier-engine system to accomplish the remainder of the descent as programed. Examples of the minimum-thrust trajectories are given in figure 13.5. For efficiency, a low minimum-thrust level is desirable. Lateral guidance is initiated during the minimum-thrust phase and continued for the approximate remainder of the descent. The pitch-and-yaw components of measured velocity are fed to the corresponding gyros for control of the vehicle attitude and for nulling of the lateral velocity. This is accomplished by "tipping" the spacecraft thrust axis to achieve the necessary velocity decrements in the pitch-and-yaw directions. Thus, when no lateral velocity component exists, the spacecraft thrust axis will be vertically alined, lateral velocity will be zero within tolerable limits, and only the vertical component of velocity will remain.

When the point $(h_t, \dot{h}_t)$ for maximum effort has been reached, as indicated from sensed velocity and altitude, the second mode begins, calling for full vernier thrust, and, at the same time, closes the guidance loop for control of thrust, as a function of measured altitude and the vertical (roll-axis) component of velocity.

The spacecraft will descend in this mode until a specified point [2] $(h_f, \dot{h}_f)$ is reached just off the lunar surface, from which the spacecraft will descend either in "free fall" or at constant velocity (1 lunar g thrust), to the lunar surface. The "target" values of h_f and $\dot{h}_f$ will depend upon the accuracy of the h, $\dot{h}$ measurements and, perhaps, upon other landing restrictions as well. Whatever the nature of this final descent to touchdown, it is important to minimize rotational rates of the spacecraft (which might be caused by shutdown transients of the vernier engines) so as not to impair the stability at landing.

Once on the surface, electrical power is removed from the guidance equipment, and the spacecraft is ready to initiate its postlanding operating sequence, beginning, it is hoped, with a report of physical survival to Earth.

[2] At this point, and at a proportionate expenditure in fuel, a "hovering" mode could be introduced for final landing-site selection via real-time TV to Earth. Lateral excursions of the spacecraft are readily accomplished directly with the lateral guidance mode described.

CHAPTER 14

Planetary Approach Guidance

TABLE 11.1 SHOWS that the miss distance at Mars or Venus resulting from Earth-based midcourse guidance is several thousand kilometers. For many missions, such as landing in a given area on the surface or passing by the planet at a controlled distance, such a target dispersion will be unacceptable. Since the capabilities for guidance from the Earth are exhausted by Earth-based midcourse guidance, further guidance must be contained in the vehicle. We shall use the term "approach guidance" to describe guidance performed in the region, for example, from 100 000 to several million kilometers from the planet. Its purpose would be to reduce the miss distance from that remaining after midcourse guidance to a value acceptable for the mission. Guidance still closer to the planet is usually called terminal guidance and will not be discussed here (see ref. 12).

In the following paragraphs, all quantities will be referred to the target planet. Note that a ballistic approach to the planet must be hyperbolic. Also, in the region of approach guidance, the speed of the spacecraft will be almost constant, and its path will be very nearly a straight line along the incoming asymptote of the hyperbola.

We first observe that fuel economy demands that the guidance be performed as far from the planet as possible. For the assumption of straight-line motion, the effect of a lateral velocity maneuver is $d=r\Delta V/v_h$, where ΔV is the magnitude of the maneuver, r is the distance from the planet at which the maneuver is executed, v_h is the speed of the spacecraft, and d is the effect, at the target, of the maneuver. (The effects of target focusing are ignored here.) Thus, if we have to correct a miss of 5000 kilometers, and if the approach speed is 5 km/sec (a typical value), and if we wish to perform the correction at 10^6 kilometers, then a maneuver of $\Delta V=25$ m/sec is needed. If we wish to make the maneuver at $r=10^5$ km, we shall require 250 m/sec. Note that for a rocket propellant having a specific impulse of 300 seconds, 1 percent of the spacecraft weight in fuel yields a maneuver of 30 m/sec.

We next inquire as to what measurements can be made from the spacecraft to compute the required maneuvers. Clearly, some measurements will have to be made with respect to the planet. Radar

measurements from 10^6 kilometers require excessive power, but optical measurements seem entirely feasible. At $r = 10^6$ kilometers, we should have no difficulty in identifying the planet. Since optical measurements generally yield angular information, we shall assume, as shown in figure 14.1, that the angles, as seen from the spacecraft, between the planet and suitably selected stars are measured. Figure 14.1 shows the use of the two stars in the direction of the x- and y-axes. The nominal path of the spacecraft is along the z-axis.

The measurements shown in figure 14.1 are adequate to determine miss distance at the planet, but they do not determine the distance r of the spacecraft from the planet. The quantity r will be needed in order to determine the time at which observations are to be taken and maneuvers made; r can be found either by (1) measuring, in the spacecraft, the angular diameter of the planet; (2) making radar measure-

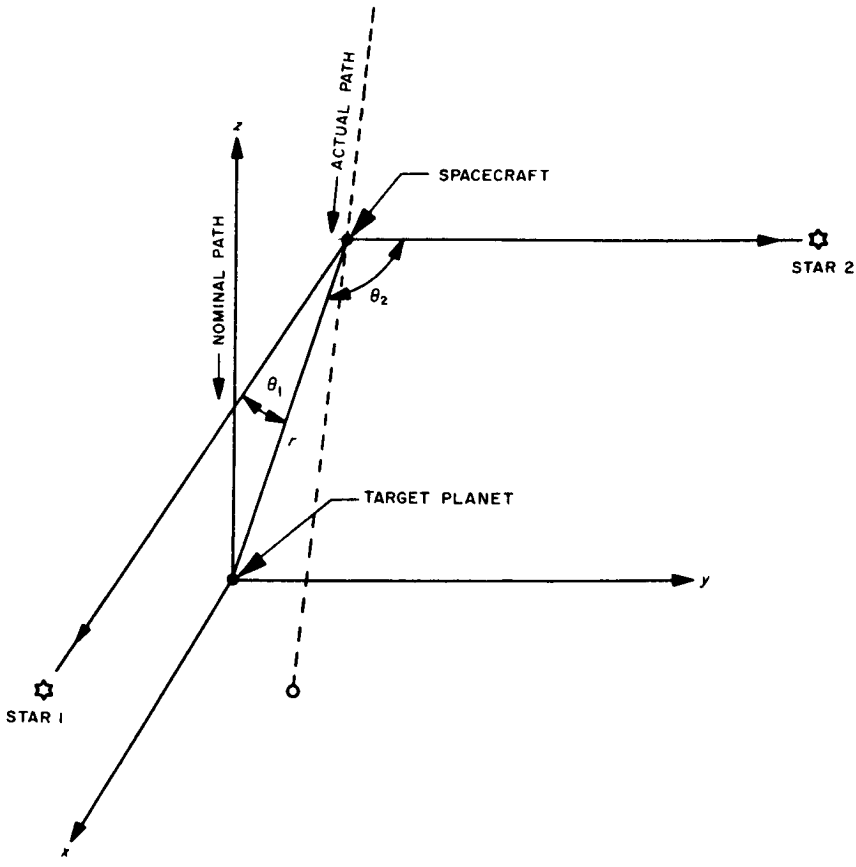


FIGURE 14.1—Use of two stars in direction of x- and y-axes.

ments, from the Earth, of range to the spacecraft; or (3) a combination of (1) and (2).

Figures 14.2 and 14.3 show how accurately the angular measurements of figure 14.1 determine the miss distance. The accuracy has been normalized with respect to the rms angular error in the observation of the angles and with respect to the greatest distance at which an observation is made. For example, suppose two measurements of the angles θ_1 and θ_2 are made at distances of 10^6 and 5×10^5 km, with an accuracy of 10^{-3} rad. From the upper curve of figure 14.2, we read, for $\rho=0.5$, $D=2.0$. Multiplying $2.0\times10^6\times10^{-3}$ yields an accuracy of 2000 kilometers, meaning that we can predict where the spacecraft will land with an rms accuracy of 2000 kilometers. In the lower curve of figure 14.2, five measurements are assumed: one at $\rho=1$, one at the value plotted, and the remaining three equally spaced; i.e., the value plotted for $\rho=0.6$ ($D=2.6$) assumes that measurements are made at $\rho=0.6$, 0.7, 0.8, 0.9, and 1.0.

Figure 14.2 assumes that no a priori data are used. However, some information will be available concerning the trajectory; in particular, we shall know something (statistically) about the direction from which the spacecraft is coming. Figure 14.3 shows the effect of such

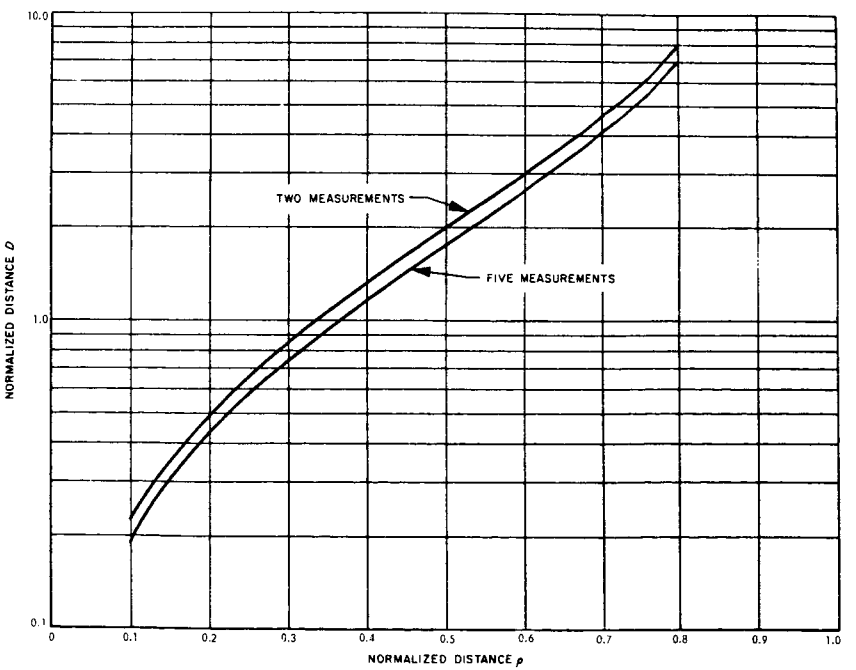


FIGURE 14.2—Normalized approach guidance accuracy.

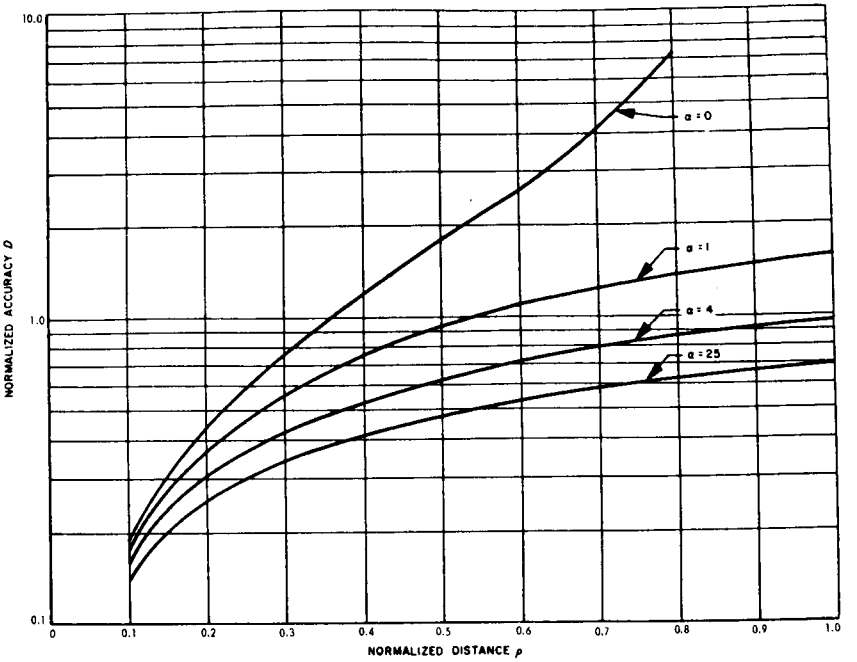


FIGURE 14.3—Approach guidance accuracy with a priori data.

a priori data. The quantity $\alpha=(\sigma_s/\sigma_a)^2$, where σ_s is the rms error in θ_1 and θ_2, and σ_a is the rms error in our a priori knowledge of the direction cosines of the incoming asymptote. As one would expect, large α, corresponding to more precise a priori information, yields greater accuracy.

To construct a practical example, suppose that the sensor accuracy is 10^{-4} rad rms, $\alpha=4$, and five measurements are made from $r=2\times10^6$ km to $r=250$ kilometers. From figure 14.3, for $\rho=0.125$, we read $D=0.2$. Thus, the accuracy of our knowledge of the trajectory is 40 kilometers rms.

Further analyses would be needed to determine the optimum distance at which to make maneuvers, the optimum number of maneuvers, the effects of errors in executing maneuvers, and so on. In general, it will be found effective to perform more than one maneuver.

CHAPTER 15

Capsule Control

THE TERM "CAPSULE" as used here implies a separate body carried by the spacecraft which is released somewhere near the target planet. The mission of the capsule is to land and survive on the surface of the planet until useful experiments have been performed and the results telemetered back to Earth.

There are several phases to the capsule control problem, and they can best be listed in chronological order. Following separation from the spacecraft, the capsule moves away at a constant velocity (typically 1 to 3 ft/sec). During this coasting phase, attitude control is usually required. The problem of capsule attitude control is, in many ways, similar to the spacecraft attitude control problem. When the distance between the spacecraft and the capsule is large enough, rocket motors on the capsule are fired to deflect the capsule from the spacecraft flyby trajectory to one which will impact the planet. Commanded turns may or may not be required for this maneuver. Next, the capsule coasts until entering the planetary atmosphere. Control may be required to orient the capsule immediately before entry. After entry the capsule is slowed down during its descent to a velocity which will allow a satisfactory landing. Finally, stabilization, erection, and other landed operations are performed.

CRUISE AND DEFLECTION

For a capsule, unlike the spacecraft, there is a reasonable possibility that only two-axis control will be required. This gives rise to the possibility of using spin stabilization instead of an active system. For this type of control, the capsule is spun immediately after separation from the spacecraft and remains spinning during the firing of the deflection motor. Since the spinning stabilizes the capsule during the cruise phase between separation and deflection, the capsule is separated in the orientation required during the motor burn phase. Spinning also stabilizes the capsule during the deflection maneuver, obviating the requirement for an autopilot. Sometime before entry into the planetary atmosphere, the capsule is despun. This is required to achieve proper dynamics during the entry phase.

238-635 O—67——10

Even though this technique of capsule attitude control is very simple, there are some errors associated with it. Between separation and spinup, the capsule is tumbling at rates that are estimated to be up to 1 deg/sec. Thus, if spinup does not occur within the first several seconds after separation, a significant pointing error will result. This initial rate gives rise to another error. It causes the spin axis of the spacecraft to nutate after spinup has occurred. This can cause errors of the order of 1°. Another error arises because of torque misalinements. Ideally, the spinup thrustors would be alined with the principal moments of inertia of the capsule. Since in practice it is very difficult to locate these axes in precisely the designated orientation, torque misalinements occur. These error torques again cause a rotating motion. Finally, in some capsule configurations it may not be possible to design the maximum moment of inertia about the spin axis. This condition is necessary, however, in order to have a stable spinning body.

If spin stabilization is not used for the attitude-control scheme, the techniques mentioned in "Attitude Control" (ch. 9) can be adapted to this use. In some cases it may be desirable to use a complete three-axis system. The final choice will be determined by the mission requirements.

ENTRY

The entry phase occurs when the capsule enters the planetary atmosphere. Assuming a ballistic entry (aerodynamic braking with high deceleration), the capsule encounters forces large enough to cause decelerations as high as 150 to 200 g. This is the phase during which a heat shield is required to protect the capsule from the high temperatures that occur.

In capsule designs being considered, the shape is chosen to be aerodynamically stable. The only control considered occurs before entry and consists of keeping the angle of attack of the heat shield small enough so that shielding can be reduced on the sides of the capsule. To maintain active control during the high deceleration portion of this phase would require actuators with a very high torquing capability, which are impractical.

DESCENT

This phase occurs between entry and touchdown on the planet. The primary use for control during this phase is to slow down the capsule to a velocity at which it can land and survive.

One technique for achieving this goal is the use of parachutes. There may be one or more stages to such a system. For a Mars

landing, one of the possible problems involved is that of surface winds. Estimates indicate that velocities up to 200 ft/sec may occur. If the capsule experiments are not capable of accepting the resulting shock that would occur at impact from this velocity, another descent scheme may be necessary.

For a truly soft landing, a completely controlled descent by means of retrorockets may be required. Such a system is used on the Surveyor spacecraft for a soft lunar landing. The disadvantage of this system is that it is considerably more complex than the use of parachutes. A complete analysis is required to show the tradeoffs involved in choosing between these two systems.

LANDING OPERATIONS

After the capsule has successfully landed on the surface of the planet, it must right itself into a workable position, and various equipment must be erected. Antennas have to be oriented in the proper direction for efficient communications. It may be necessary to scan the surface with TV cameras. Conventional articulation-control systems can be used to satisfy these requirements.

If it is necessary to establish a horizontal platform and determine the latitude or the direction of north, several techniques are available. An accelerometer could be used to determine the horizontal direction. Gyros could be used to find North and also to establish the latitude of the landing site.

Generally speaking, the control problems encountered in a landed capsule do not appear to be formidable. With careful analysis and design, adequate and reliable systems seem to be within today's state of the art.

References

1. SCULL, J. R.: Guidance of Space Vehicles. Navigation, vol. 8, no. 1, Spring, 1961, pp. 24–33.
2. SCULL, J. R.: The Application of Optical Sensors for Lunar and Planetary Space Vehicles. Heat and Light Sensing AGARDograph 71, ch. 25, Pergamon Press, Bristol, England, 1961.
3. DRAPER, C. S.; WRIGLEY, WALTER; AND GROHE, L. R.: The Floating Integrating Gyro and Its Application to Geometrical Stabilization Problems on Moving Bases. MIT paper, I.A.S., S.M.F. Fund Paper No. FF–13, New York, Jan. 1955.
4. WRIGLEY, WALTER; WOODBURY, R. B.; AND HOVORKA, JOHN: Inertial Guidance. MIT paper, I.A.S., S.M.F. Fund Paper No. FF–16, New York, Jan. 1957.
5. KLASS, PHILIP: Inertial Guidance. Aviation Week, Special Report, McGraw-Hill Book Co., Inc., 1956.
6. PITMAN, G. R.; ET AL.: Inertial Guidance. John Wiley & Sons, 1962.
7. LEONDES, C. T.; ET AL.: Guidance and Control of Aerospace Vehicles. McGraw-Hill Book Co., Inc., 1963.
8. SAVANT, C. J.: Basic Feedback Control System Design. McGraw-Hill Book Co., Inc., 1958.
9. GOODE, H. H.; AND MACHOL, R. E.: System Engineering. McGraw-Hill Book Co., Inc., 1957.
10. SMITH, A. H.; FREDRICKSON, C. D.; AND LEVENTHAL, E. L.: Internal Communication. Jet Propulsion Laboratory, California Institute of Technology, 1962.
11. MORSE, J. G.: Space Power Systems. Lecture Notes for UCLA Short Course X494MN, sec. IV, F–2—Radioisotope Fueled Power Supplies, 1964.
12. GATES, C. R.; SCULL, J. R.; AND WATKINS, K. S.: Space Guidance. Astronautics, American Rocket Society, Nov. 1961 (also revised and published as ARS–2113–61).

U.S. GOVERNMENT PRINTING OFFICE : 1967—O-238-635

"The aeronautical and space activities of the United States shall be conducted so as to contribute . . . to the expansion of human knowledge of phenomena in the atmosphere and space. The Administration shall provide for the widest practicable and appropriate dissemination of information concerning its activities and the results thereof."

—NATIONAL AERONAUTICS AND SPACE ACT OF 1958

NASA SCIENTIFIC AND TECHNICAL PUBLICATIONS

TECHNICAL REPORTS: Scientific and technical information considered important, complete, and a lasting contribution to existing knowledge.

TECHNICAL NOTES: Information less broad in scope but nevertheless of importance as a contribution to existing knowledge.

TECHNICAL MEMORANDUMS: Information receiving limited distribution because of preliminary data, security classification, or other reasons.

CONTRACTOR REPORTS: Scientific and technical information generated under a NASA contract or grant and considered an important contribution to existing knowledge.

TECHNICAL TRANSLATIONS: Information published in a foreign language considered to merit NASA distribution in English.

SPECIAL PUBLICATIONS: Information derived from or of value to NASA activities. Publications include conference proceedings, monographs, data compilations, handbooks, sourcebooks, and special bibliographies.

TECHNOLOGY UTILIZATION PUBLICATIONS: Information on technology used by NASA that may be of particular interest in commercial and other non-aerospace applications. Publications include Tech Briefs, Technology Utilization Reports and Notes, and Technology Surveys.

Details on the availability of these publications may be obtained from:

SCIENTIFIC AND TECHNICAL INFORMATION DIVISION

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

Washington, D.C. 20546